POLYNUCLEOTIDES AND AMINOACID SEQUENCES FROM STAPHYLOCOCCUS AUREUS

Publication number: WO9731114

Publication date: 1997-08-28

Inventor: BURNHAM MARTIN KARL RUSSELL (US); HODGSON JOHN EDWARD (US)

Applicant: SMITHKLINE BEECHAM PLC (GB); BURNHAM MARTIN KARL RUSSELL (US); HODGSON JOHN EDWARD (US)

Classification:

- International: ***G01N33/50; A61K38/00; A61K39/085; A61K39/395; A61K48/00; A61P31/04; C07K14/31; C07K16/12; C12N1/19; C12N1/21; C12N15/00; C12N15/09; C12N15/31; C12P21/02; C12P21/08; G01N33/15;*** *A61K39/00; C12R1/445;* **G01N33/50; A61K38/00; A61K39/085; A61K39/395; A61K48/00; A61P31/00; C07K14/195; C07K16/12; C12N1/19; C12N1/21; C12N15/00; C12N15/09; C12N15/31; C12P21/02; C12P21/08; G01N33/15;** A61K39/00; (IPC1-7): C12N15/31; A61K39/085; A61K48/00; C07K14/31; C07K16/12; C12Q1/68; G01N33/569; G01N33/68

- European: C07K14/31

Application number: WO1997GB00524 19970225

Priority number(s): GB19960004045 19960226

Also published as:

WO9731114 (A3)
EP0822987 (A3)
EP0822987 (A2)
EP0822987 (A0)

Cited documents:

WO9422895
WO9400493
XP002034046
XP002034047
XP000676708
XP002034045
less <<

Report a data error here

Abstract of **WO9731114**

This invention relates to Staphylococcal polynucleotides, polypeptides encoded by such polynucleotides, the uses of such polynucleotides and polypeptides, as well as the production of such polynucleotides and polypeptides and recombinant host cells transformed with the polynucleotides. This invention also relates to inhibiting the biosynthesis or action of such polynucleotides or polypeptides and to the use of such inhibitors in therapy.

Family list
11 family members for: **WO9731114**
Derived from 9 applications

Back to WO973

1 **NOVEL SPO-REL**
Inventor: BURNHAM MARTIN KARL RUSSEL (US); GENTRY DANIEL ROBERT (US); (+2)
Applicant: SMITHKLINE BEECHAM PLC (GB); SMITHKLINE BEECHAM CORP (US)
EC: C07K14/31
IPC: ***G01N33/53; A61K38/00; A61K39/00*** (+53)
Publication info: CA2214866 A1 - 1998-04-29

2 **POLYNUCLEOTIDES AND AMINOACID SEQUENCES FROM STAPHYLOCOCCUS AUREUS**
Inventor: BURNHAM MARTIN K R (US); HODGSON JOHN EDWARD (US)
Applicant: SMITHKLINE BEECHAM PLC (GB)
EC: C07K14/31
IPC: ***G01N33/50; A61K38/00; A61K39/085*** (+40)
Publication info: EP0822987 A2 - 1998-02-11

3 **Spo-rel, a Staphylococcus relA/spot homologue**
Inventor: HODGSON JOHN EDWARD (US); BURNHAM MARTIN K R (US); (+2)
Applicant: SMITHKLINE BEECHAM CORP (US); SMITHKLINE BEECHAM PLC (GB)
EC: C07K14/31
IPC: ***G01N33/53; A61K38/00; A61K39/00*** (+48)
Publication info: EP0839910 A2 - 1998-05-06
EP0839910 A3 - 2000-01-19

4 **No title available**
Inventor:
Applicant:
EC:
IPC:
Publication info: GB9604045D D0 - 1996-04-24

5 **NEW SPO-REL**
Inventor: HODGSON JOHN EDWARD; BURNHAM MARTIN KARL R; (+2)
Applicant: SMITHKLINE BEECHAM CORP; SMITHKLINE BEECHAM PLC
EC: C07K14/31
IPC: ***G01N33/53; A61K38/00; A61K39/00*** (+61)
Publication info: JP10225296 A - 1998-08-25

6 **POLYNUCLEOTIDES AND AMINOACID SEQUENCES FROM STAPHYLOCOCCUS AUREUS**
Inventor:
Applicant:
EC: C07K14/31
IPC: ***G01N33/50; A61K38/00; A61K39/085*** (+47)
Publication info: JP11506022T T - 1999-06-02

7 **DNA encoding spo-rel polypeptides**
Inventor: BURNHAM MARTIN KARL RUSSEL (US); GENTRY DANIEL ROBERT (US); (+2)
Applicant: SMITHKLINE BEECHAM CORP (US); SMITHKLINE BEECHAM PLC (GB)
EC: C07K14/31
IPC: ***G01N33/53; A61K38/00; A61K39/00*** (+47)
Publication info: US5989864 A - 1999-11-23

8 **Spo-rel**
Inventor: BURNHAM MARTIN KARL RUSSEL (US); GENTRY DANIEL ROBERT (US); (+2)
Applicant:
EC: C07K14/31
IPC: ***A61P31/04; C07K14/31***; *A61K38/00* (+10)
Publication info: US6365159 B1 - 2002-04-02

9 **POLYNUCLEOTIDES AND AMINOACID SEQUENCES FROM STAPHYLOCOCCUS AUREUS**
Inventor: BURNHAM MARTIN KARL RUSSELL (US); HODGSON JOHN EDWARD (US)
Applicant: SMITHKLINE BEECHAM PLC (GB); BURNHAM MARTIN KARL RUSSELL (US); (+1)
EC: C07K14/31
IPC: ***G01N33/50; A61K38/00; A61K39/085*** (+40)

Publication info: WO9731114 A2 - 1997-08-28
WO9731114 A3 - 1997-10-23

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/31, C07K 14/31, 16/12, A61K 39/085, 48/00, G01N 33/68, 33/569, C12Q 1/68

A2

(11) International Publication Number: WO 97/31114

(43) International Publication Date: 28 August 1997 (28.08.97)

(21) International Application Number: PCT/GB97/00524

(22) International Filing Date: 25 February 1997 (25.02.97)

(30) Priority Data:
9604045.6 26 February 1996 (26.02.96) GB

(71) Applicant *(for all designated States except US)*: SMITHKLINE BEECHAM PLC [GB/GB]; New Horizons Court, Brentford, Middlesex TW8 9EP (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: BURNHAM, Martin, Karl, Russell [GB/US]; SmithKline Beecham Pharmaceuticals, 1250 South Collegeville Road, P.O. Box 5089, Collegeville, PA 19426-0989 (US). HODGSON, John, Edward [GB/US]; SmithKline Beecham Pharmaceuticals, 1250 South Collegeville Road, P.O. Box 5089, Collegeville, PA 19426-0989 (US).

(74) Agent: GIDDINGS, Peter, John; SmithKline Beecham, Corporate Intellectual Property, Two New Horizons Court, Brentford, Middlesex TW8 9EP (GB).

(81) Designated States: JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: POLYNUCLEOTIDES AND AMINOACID SEQUENCES FROM STAPHYLOCOCCUS AUREUS

(57) Abstract

This invention relates to Staphylococcal polynucleotides, polypeptides encoded by such polynucleotides, the uses of such polynucleotides and polypeptides, as well as the production of such polynucleotides and polypeptides and recombinant host cells transformed with the polynucleotides. This invention also relates to inhibiting the biosynthesis or action of such polynucleotides or polypeptides and to the use of such inhibitors in therapy.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

POLYNUCLEOTIDES AND AMINOACID SEQUENCES FROM STAPHYLOCOCCUS AUREUS

FIELD OF THE INVENTION

This invention relates to newly identified polynucleotides, polypeptides encoded by such polynucleotides, the use of such polynucleotides and polypeptides, as well as the production of such polynucleotides and polypeptides and recombinant host cells transformed with the polynucleotides. This invention also relates to inhibiting the biosynthesis or action of such polypeptides and to the use of inhibitors in therapy.

BACKGROUND OF THE INVENTION

The Staphylococci make up a medically important genera of microbes. They are known to produce two types of disease, invasive and toxigenic. Invasive infections are characterized generally by abscess formation effecting both skin surfaces and deep tissues. *Staphlococcus aureus* is the second leading cause of bacteremia in cancer patients. Osteomyelitis, septic arthritis, septic thrombophlebitis and acute bacterial endocarditis are also relatively common. There are at least three clinical conditions resulting from the toxigenic properties of Staphylococci. The manifestation of these diseases result from the actions of exotoxins as opposed to tissue invasion and bacteremia. These conditions include: Staphylococcal food poisoning, scalded skin syndrome and toxic shock syndrome.

While certain Staphylococcal proteins associated with pathogenicity have been identified, e.g., coagulase, hemolysins, leucocidins and exo and enterotoxins, very little is known concerning the temporal expression of genes of bacterial pathogens during infection and disease progression in a mammalian host. Discovering the sets of genes the bacterium is likely to be expressing at the different stages of infection, particularly when an infection is established, provides critical information for the screening and characterization of novel antibacterials which can interrupt pathogenesis, by identifying possible previously unrecognised targets.

Recently several novel approaches have been described which purport to follow global gene expression during infection (Chuang, S. et al. [1993] Global Regulation of Gene Expression in *Escherichia coli* J. Bacteriol. 175, 2026-2036, Mahan, M.J. et al. [1993] Selection of Bacterial Virulence Genes That Are Specifically Induced in Host Tissues SCIENCE 259, 686-688, Hensel, M. et al. [1995] Simultaneous Identification of Bacterial Virulence Genes by Negative Selection SCIENCE 269, 400-403). These new techniques have so far been demonstrated with gram negative pathogen infections and not with infections with gram positives presumably due to the much slower development of

global transposon mutagenesis and suitable vectors needed for these strategies in these organisms, and in the case of that process described by Chuang, S. et al.[1993] the difficulty of isolating suitable quantities of bacterial RNA free of mammalian RNA derived from the infected tissue to furnish bacterial RNA labelled to sufficiently high specific activity. The present invention employs a novel technology to determine gene expression in the pathogen at different stages of infection of the mammalian host.

DETAILED DESCRIPTION OF THE INVENTION

A novel aspect of this invention is the use of a suitably labelled oligonucleotide probe which anneals specifically to the bacterial ribosomal RNA in Northern blots of bacterial RNA preparations from infected tissue. Using the more abundant ribosomal RNA as a hybridisation target greatly facilitates the optimisation of a protocol to purify bacterial RNA of a suitable size and quantity for RT-PCR from infected tissue. Techniques reported in the scientific literature which are of use in purifying *Staphylococcus aureus* RNA from bacteria grown *in vitro* are unsuccessful when applied to infected tissue.

In a first aspect therefore, the invention provides a method of identifying genes transcribed in an organism in infected host tissue by identifying mRNA present using RT-PCR, characterised in that a bacterial mRNA preparation is obtained from total RNA from infected tissue by enriching for bacterial RNA by a suitable bacterial disruption technique in order to selectively damage mammalian RNA and at the same time give sufficient quantities of bacterial RNA for RT-PCR, and wherein the conditions for selectively enriching for bacterial RNA are determined by probing with an oligonucleotide probe specific to bacterial ribosomal RNA.

This process of optimisation preferably uses a unique labelled oligonucleotide probe to bacterial ribosomal RNA which is used in Northern experiments against the experimental RNA preparations to determine those conditions which give optimal levels of bacterial RNA. As bacterial ribosomal RNA is present at 2-4 orders of magnitude in amount to bacterial mRNA species this detection procedure provides a suitably sensitive indication to the existence and quantity of bacterial RNA in the presence of the vastly greater levels of mammalian RNA from the infected tissue. This detection system may be used in conjunction with the visualisation of total RNA by ethidium bromide staining of 1% agarose gels on which it has been run out. On these gels mammalian ribosomal RNA migrates at a different rate to bacterial ribosomal RNA and so can be identified. Surprisingly, those disruption conditions which were found to just lead to the loss of

mammalian RNA gave the best preparations of bacterial RNA as judged by the Northern experiment. A suitable oligonucleotide useful for applying this method to genes expressed in *Staphylococcus aureus* is 5'-gctcctaaaaggttactccaccggc-3' [SEQ ID NO:91].

Use of the technology of the present invention enables identification of bacterial genes transcribed during infection, inhibitors of which would have utility in anti-bacterial therapy. Specific inhibitors of such gene transcription or of the subsequent translation of the resultant mRNA or of the function of the corresponding expressed proteins would have utility in anti-bacterial therapy

The present invention provides a polynucleotide having the DNA sequence given in any of sequences set forth in, or selected from the group consisting essentially of, SEQUENCE 1 [SEQ ID Nos:1,4,7,10,13,16,19,22,25,28, 31,34,37,40,43,46,49,52,55,58,61,64,67,70,73,76] of Table 1, or any combination of the sequences thereof. The invention further provides a polynucleotide encoding a protein from *S. aureus* WCUH 29 and characterized in that it comprises the DNA sequence given in any of sequences set forth in SEQUENCE 1 [SEQ ID Nos: 1,4,7,10,13,16,19,22,25,28,31,34,37,40,43,46,49,52,55,58,61,64, 67,70,73,76] of Table 1. The polynucleotides having the DNA sequence given in each sequence set forth in SEQUENCE 1 [SEQ ID Nos: 1,4,7,10,13,16,19,22,25,28, 31,34,37,40,43,46,49,52,55,58,61,64,67,70,73,76] of Table 1 were obtained from the sequencing of a library of clones of chromosomal DNA of *S.aureus* WCUH 29 in *E.coli*.

S. aureus WCUH 29 has been deposited at the National Collection of Industrial and Marine Bacteria Ltd. (NCIMB), Aberdeen, Scotland under number NCIMB 40771 on 11 September 1995.

The present invention also provides a novel protein from *Staphylococcus. aureus* WCUH29 obtainable by expression of a gene characterised in that it comprises the DNA sequence given in any of sequences set forth in SEQUENCE 1 [SEQ ID Nos: 1,4,7,10, 13,16,19,22,25,28,31,34,37,40,43,46,49,52,55,58,61,64,67,70,73,76] of Table 1, or a fragment, analogue or derivative thereof.

The present invention further relates to a novel protein from *Staphylococcus. aureus* WCUH29, characterised in that it comprises the amino acid sequence given in any of the sequences set forth in, or selected from the group consisting essentially of,

SEQUENCE 2 [SEQ ID Nos: 79,80,81,82,83,84,85,86,87,88,89,90] of Table 1, or a fragment, analogue or derivative thereof.

The invention also relates to a polypeptide fragment of the protein, having the amino acid sequence given in any of the sequences set forth in SEQUENCE 2 [SEQ ID Nos: 79,80,81.82,83,84,85,86,87,88,89,90] of Table 1, or a derivative thereof.

Hereinafter the term polypeptide(s) will be used to refer to the protein and its fragments, analogues or derivatives.

In accordance with another aspect of the present invention, there are provided polynucleotides (DNA or RNA) which encode such polypeptides.

The invention also relates to novel oligonucleotides, including the sequences set forth in SEQUENCE 3 [SEQ ID Nos: 2,5,8,11,14,17,20,23,26,29,32,35,38,41,44,47, 50,53,56,59,62,65,68,71,74,77] and 4 [SEQ ID Nos: 3,6,9,12,15,18,21,24,27,30, 33,36,39,42,45,48,51,54,57,60.63,66,69,72,75,78] of Table 1, derived from the sequences set forth in SEQUENCE 1 [SEQ ID Nos: 1,4,7,10,13,16,19,22,25,28,31,34, 37,40,43,46,49,52,55,58,61,64,67,70,73,76] of Table 1 which can act as PCR primers in the process herein described to determine whether or not the *Staphylococcus aureus* genes identified herein in whole or in part are transcribed in infected tissue. It is recognised that such sequences will also have utility in diagnosis of the stage of infection and type of infection the pathogen has attained.

Each of the DNA sequences provided herein may be used in the discovery and development of antibacterial compounds. The encoded protein upon expression can be used as a target for the screening of antibacterial drugs. Additionally, the DNA sequences encoding regions of the encoded protein or Shine-Delgarno or other translation facilitating sequences of the respective mRNA can be used to construct antisense sequences to control the expression of the coding sequence of interest. Furthermore, many of the sequences disclosed herein also provide regions upstream and downstream from the encoding sequence. These sequences are useful as a source of regulatory elements for the control of bacterial gene expression. Such sequences are conveniently isolated by restriction enzyme action or synthesized chemically and introduced, for example, into promoter identification strains. These strains contain a reporter structural gene sequence located downstream from a restriction site such that if an active promoter is inserted, the reporter gene will be expressed.

Although each of the sequences may be employed as described above, this invention also provides several means for identifying particularly useful target genes. The first of these approaches entails searching appropriate databases for sequence matches. Thus, if a homologue exists, the Staphylococcal-like form of this gene would likely play an analogous role. For example, a Staphylococcal protein identified as homologous to a cell surface protein in another organism would be useful as a vaccine candidate. To the extent such homologies have been identified for the sequences disclosed herein they are reported along with the encoding sequence.

To obtain the polynucleotide encoding the protein using any DNA sequence given in a SEQ ID NO 1 typically a library of clones of chromosomal DNA of *S.aureus* WCUH 29 in *E.coli* or some other suitable host is probed with a radiolabelled oligonucleotide, preferably a 17mer or longer, derived from the partial sequence. Clones carrying DNA identical to that of the probe can then be distinguished using high stringency washes. By sequencing the individual clones thus identified with sequencing primers designed from the original sequence it is then possible to extend the sequence in both directions to determine the full gene sequence. Conveniently such sequencing is performed using denatured double stranded DNA prepared from a plasmid clone. Suitable techniques are described by Maniatis, T., Fritsch, E.F. and Sambrook, J. in MOLECULAR CLONING, A Laboratory Manual [2nd edition 1989 Cold Spring Harbor Laboratory. see Screening By Hybridization 1.90 and Sequencing Denatured Double-Stranded DNA Templates 13.70].

A polynucleotide of the present invention may be in the form of RNA or in the form of DNA, which DNA includes cDNA, genomic DNA, and synthetic DNA. The DNA may be double-stranded or single-stranded, and if single stranded may be the coding strand or non-coding (anti-sense) strand. The coding sequence which encodes the polypeptide may be identical to the coding sequence of any of the sequences of SEQUENCE 1 [SEQ ID Nos:1,4,7,10,13,16,19,22,25,28,31,34,37,40,43,46,49,52,55,58,61,64,67,70,73,76] of Table 1 or may be a different coding sequence which coding sequence, as a result of the redundancy or degeneracy of the genetic code, encodes the same polypeptide.

The present invention includes variants of the hereinabove described polynucleotides which encode fragments, analogues and derivatives of the polypeptides of the invention, and in particular polypeptides characterized by the deduced amino acid sequences set forth in each SEQUENCE 2 [SEQ ID Nos: 79,80,81,82,83,84,

85,86,87,88.89,90] of Table 1. The variant of the polynucleotide may be a naturally occurring allelic variant of the polynucleotide or a non-naturally occurring variant of the polynucleotide.

Thus, the present invention includes polynucleotides encoding the same polypeptides of the invention, and in particular characterized by the deduced amino acid sequences set forth in each SEQUENCE 2 [SEQ ID Nos: 79,80,81,82,83,84,85,86,87, 88,89,90] of Table 1 as well as variants of such polynucleotides which variants encode for a fragment, derivative or analogue of the polypeptide. Such nucleotide variants include deletion variants, substitution variants and addition or insertion variants.

The polynucleotide may have a coding sequence which is a naturally occurring allelic variant of the coding sequence characterized by the DNA sequence of any of the sequences set forth in Table 1 as SEQUENCE 1 [SEQ ID Nos:1,4,7,10,13,16,19,22,25, 28,31,34,37,40,43,46,49,52,55,58,61,64,67,70,73,76]. As known in the art, an allelic variant is an alternate form of a polynucleotide sequence which may have a substitution, deletion or addition of one or more nucleotides, which does not substantially alter the function of the encoded polypeptide.

The polynucleotide which encodes for the mature polypeptide may include only the coding sequence for the mature polypeptide or the coding sequence for the mature polypeptide and additional coding sequence such as a leader or secretory sequence or a proprotein sequence.

Thus, the term "polynucleotide encoding a polypeptide" encompasses a polynucleotide which includes only coding sequence for the polypeptide as well as a polynucleotide which includes additional coding and/or non-coding sequence.

The present invention therefore includes polynucleotides, wherein the coding sequence for the mature polypeptide may be fused in the same reading frame to a polynucleotide sequence which aids in expression and secretion of a polypeptide from a host cell, for example, a leader sequence which functions as a secretory sequence for controlling transport of a polypeptide from the cell. The polypeptide having a leader sequence is a preprotein and may have the leader sequence cleaved by the host cell to form the mature form of the polypeptide. The polynucleotides may also encode for a proprotein which is the mature protein plus additional 5' amino acid residues. A mature protein having a prosequence is a proprotein and is an inactive form of the protein. Once the prosequence is cleaved an active mature protein remains.

Thus, for example, the polynucleotide of the present invention may encode for a mature protein, or for a protein having a prosequence or for a protein having both a prosequence and a presequence (leader sequence). Further, the amino acid sequences provided herein show a methionine residue at the NH_2-terminus. It is appreciated, however, that during post-translational modification of the peptide, this residue may be deleted. Accordingly, this invention contemplates the use of both the sequences.

An expression vector is constructed so that the particular coding sequence is located in the vector with the appropriate regulatory sequences, the positioning and orientation of the coding sequence with respect to the control sequences being such that the coding sequence is transcribed under the "control" of the control sequences (i.e., RNA polymerase which binds to the DNA molecule at the control sequences transcribes the coding sequence). Modification of the coding sequences may be desirable to achieve this end. For example, in some cases it may be necessary to modify the sequence so that it may be attached to the control sequences with the appropriate orientation; i.e., to maintain the reading frame. The control sequences and other regulatory sequences may be ligated to the coding sequence prior to insertion into a vector, such as the cloning vectors described above. Alternatively, the coding sequence can be cloned directly into an expression vector which already contains the control sequences and an appropriate restriction site.

Generally, recombinant expression vectors will include origins of replication and selectable markers permitting transformation of the host cell, e.g., the ampicillin resistance gene of *E. coli* and *S. cerevisiae* TRP1 gene, and a promoter derived from a highly-expressed gene to direct transcription of a downstream structural sequence. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably, a leader sequence capable of directing secretion of translated protein into the periplasmic space or extracellular medium. Optionally, the heterologous sequence can encode a fusion protein including an N-terminal identification peptide imparting desired characteristics, e.g., stabilization or simplified purification of expressed recombinant product.

The vector containing the appropriate DNA sequence as hereinabove described, as well as an appropriate promoter or control sequence, may be employed to transform an appropriate host to permit the host to express the protein.

More particularly, the present invention also includes recombinant constructs comprising one or more of the sequences as broadly described above. The constructs

comprise a vector, such as a plasmid or viral vector, into which a sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available. The following vectors are provided by way of example. Bacterial: pET-3 vectors (Stratagene), pQE70, pQE60, pQE-9 (Qiagen), pbs, pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16a, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia). Eukaryotic: pBlueBacIII (Invitrogen), pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene) pSVK3, pBPV, pMSG, pSVL (Pharmacia). However, any other plasmid or vector may be used as long as they are replicable and viable in the host.

Examples of recombinant DNA vectors for cloning and host cells which they can transform include the bacteriophage λ (*E. coli*), pBR322 (*E. coli*), pACYC177 (*E. coli*), pKT230 (gram-negative bacteria), pGV1106 (gram-negative bacteria), pLAFR1 (gram-negative bacteria), pME290 (non-*E. coli* gram-negative bacteria), pHV14 (*E. coli* and *Bacillus subtilis*), pBD9 (*Bacillus*), pIJ61 (*Streptomyces*), pUC6 (*Streptomyces*), YIp5 (*Saccharomyces*), a baculovirus insect cell system, YCp19 (*Saccharomyces*). See, generally, "DNA Cloning": Vols. I & II, Glover *et al.* ed. IRL Press Oxford (1985) (1987) and; T. Maniatis *et al.* ("Molecular Cloning" Cold Spring Harbor Laboratory (1982). methionine-containing and the methionineless amino terminal variants of each protein disclosed herein.

The polynucleotides of the present invention may also have the coding sequence fused in frame to a marker sequence at either the 5' or 3' terminus of the gene which allows for purification of the polypeptide of the present invention. The marker sequence may be a hexa-histidine tag supplied by the pQE series of vectors (supplied commercially by Quiagen Inc.) to provide for purification of the polypeptide fused to the marker in the case of a bacterial host.

The present invention further relates to polynucleotides which hybridize to the hereinabove-described sequences if there is at least 50% and preferably at least 70% identity between the sequences. The present invention particularly relates to Staphylococcal polynucleotides which hybridize under stringent conditions to the hereinabove-described polynucleotides . As herein used, the term "stringent conditions"

means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences. The polynucleotides which hybridize to the hereinabove described polynucleotides in a preferred embodiment encode polypeptides which retain substantially the same biological function or activity as the polypeptide of the invention. A preferred embodiment of the invention is a polynucleotide having at least a 70%, 80%, 90% or 95% identity to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting essentially of SEQ ID Nos: 79,80,81,82,83,84,85,86,87,88 and 89, or any combination of these amino acid sequences.

The deposit referred to herein will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for purposes of Patent Procedure. These deposits are provided merely as convenience to those of skill in the art and are not an admission that a deposit is required under 35 U.S.C. §112. The sequence of the polynucleotides contained in the deposited material, as well as the amino acid sequence of the polypeptides encoded thereby, are incorporated herein by reference and are controlling in the event of any conflict with any description of sequences herein. A license may be required to make, use or sell the deposited material, and no such license is hereby granted.

The terms "fragment," "derivative" and "analogue" when referring to the polypeptide of the invention, means a polypeptide which retains essentially the same biological function or activity as such polypeptide. Thus, an analogue includes a proprotein which can be activated by cleavage of the proprotein portion to produce an active mature polypeptide.

The polypeptide of the present invention may be a recombinant polypeptide, a natural polypeptide or a synthetic polypeptide, preferably a recombinant polypeptide.

The fragment, derivative or analogue of the polypeptide of the invention may be (i) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue (preferably a conserved amino acid residue) and such substituted amino acid residue may or may not be one encoded by the genetic code, or (ii) one in which one or more of the amino acid residues includes a substituent group, or (iii) one in which the polypeptide is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or (iv) one in which the additional amino acids are fused to the polypeptide, such as a leader or secretory sequence or a sequence which is employed for purification of the polypeptide or a

proprotein sequence. Such fragments, derivatives and analogues are deemed to be within the scope of those skilled in the art from the teachings herein.

The polypeptides and polynucleotides of the present invention are preferably provided in an isolated form, and preferably are purified to homogeneity.

The term "isolated" means that the material is removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotides could be part of a vector and/or such polynucleotides or polypeptides could be part of a composition, and still be isolated in that such vector or composition is not part of its natural environment.

The present invention also relates to vectors which include polynucleotides of the present invention, host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques.

In accordance with yet a further aspect of the present invention, there is therefore provided a process for producing the polypeptide of the invention by recombinant techniques by expressing a polynucleotide encoding said polypeptide in a host and recovering the expressed product. Alternatively, the polypeptides of the invention can be synthetically produced by conventional peptide synthesizers.

Host cells are genetically engineered (transduced or transformed or transfected) with the vectors of this invention which may be, for example, a cloning vector or an expression vector. The vector may be, for example, in the form of a plasmid, a cosmid, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants or amplifying the genes. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

Suitable expression vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., bacterial plasmids; phage DNA; baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA. However, any other vector may be used as long as it is replicable and viable in the host.

The appropriate DNA sequence may be inserted into the vector by a variety of procedures. In general, the DNA sequence is inserted into an appropriate restriction endonuclease site(s) by procedures known in the art.

The DNA sequence in the expression vector is operatively linked to an appropriate expression control sequence(s) (promoter) to direct mRNA synthesis. As representative examples of such promoters, there may be mentioned: LTR or SV40 promoter, the *E. coli. lac* or *trp*, the phage lambda P_L promoter and other promoters known to control expression of genes in eukaryotic or prokaryotic cells or their viruses. The expression vector also contains a ribosome binding site for translation initiation and a transcription terminator. The vector may also include appropriate sequences for amplifying expression.

In addition, the expression vectors preferably contain one or more selectable marker genes to provide a phenotypic trait for selection of transformed host cells such as dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, or such as tetracycline or ampicillin resistance in *E. coli.*

The gene can be placed under the control of a promoter, ribosome binding site (for bacterial expression) and, optionally, an operator (collectively referred to herein as "control" elements), so that the DNA sequence encoding the desired protein is transcribed into RNA in the host cell transformed by a vector containing this expression construction. The coding sequence may or may not contain a signal peptide or leader sequence. The polypeptides of the present invention can be expressed using, for example, the *E. coli* tac promoter or the protein A gene (spa) promoter and signal sequence. Leader sequences can be removed by the bacterial host in post-translational processing. See, e.g., U.S. Patent Nos. 4,431,739; 4,425,437; 4,338,397. Promoter regions can be selected from any desired gene using CAT (chloramphenicol transferase) vectors or other vectors with selectable markers. Two appropriate vectors are PKK232-8 and PCM7. Particular named bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda P_R, P_L and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Selection of the appropriate vector and promoter is well within the level of ordinary skill in the art.

In addition to control sequences, it may be desirable to add regulatory sequences which allow for regulation of the expression of the protein sequences relative to the growth of the host cell. Regulatory sequences are known to those of skill in the art, and examples include those which cause the expression of a gene to be turned on or off in response to a

chemical or physical stimulus, including the presence of a regulatory compound. Other types of regulatory elements may also be present in the vector, for example, enhancer

In some cases, it may be desirable to add sequences which cause the secretion of the polypeptide from the host organism, with subsequent cleavage of the secretory signal.

Polypeptides can be expressed in host cells under the control of appropriate promoters. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors for use with prokaryotic and eukaryotic hosts are described by Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, N.Y., (1989), the disclosure of which is hereby incorporated by reference.

Following transformation of a suitable host strain and growth of the host strain to an appropriate cell density, the selected promoter is induced by appropriate means (e.g., temperature shift or chemical induction) and cells are cultured for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents, such methods are well known to those skilled in the art.

Depending on the expression system and host selected, the polypeptide of the present invention may be produced by growing host cells transformed by an expression vector described above under conditions whereby the polypeptide of interest is expressed. The polypeptide is then isolated from the host cells and purified. If the expression system secretes the polypeptide into growth media, the polypeptide can be purified directly from the media. If the polypeptide is not secreted, it is isolated from cell lysates or recovered from the cell membrane fraction. Where the polypeptide is localized to the cell surface, whole cells or isolated membranes can be used as an assayable source of the desired gene product. Polypeptide expressed in bacterial hosts such as *E. coli* may require isolation from inclusion bodies and refolding. Where the mature protein has a very hydrophobic region which leads to an insoluble product of overexpression, it may be desirable to express a truncated protein in which the hydrophobic region has been deleted. The selection of the appropriate growth conditions and recovery methods are within the skill of the art.

The polypeptide can be recovered and purified from recombinant cell cultures by methods including ammonium sulphate or ethanol precipitation, acid extraction, anion or

cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Protein refolding steps can be used, as necessary, in completing configuration of the mature protein. Finally, high performance liquid chromatography (HPLC) can be employed for final purification steps.

Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. Polypeptides of the invention may also include an initial methionine amino acid residue.

A "replicon" is any genetic element (e.g., plasmid, chromosome, virus) that functions as an autonomous unit of DNA replication *in vivo*; i.e., capable of replication under its own control.

A "vector" is a replicon, such as a plasmid, phage, or cosmid, to which another DNA segment may be attached so as to bring about the replication of the attached segment.

A "double-stranded DNA molecule" refers to the polymeric form of deoxyribonucleotides (bases adenine, guanine, thymine, or cytosine) in a double-stranded helix, both relaxed and supercoiled. This term refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, *inter alia*, in linear DNA molecules (e.g., restriction fragments), viruses, plasmids, and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences may be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the nontranscribed strand of DNA (i.e., the strand having the sequence homologous to the mRNA).

A DNA "coding sequence of" or a "nucleotide sequence encoding" a particular protein, is a DNA sequence which is transcribed and translated into a polypeptide when placed under the control of appropriate regulatory sequences.

A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of defining the present invention, the promoter sequence is bound at the 3' terminus by a translation start codon (e.g., ATG) of a coding sequence and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site (conveniently defined by mapping with nuclease

S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase. Eukaryotic promoters will often, but not always, contain "TATA" boxes and "CAT" boxes. Prokaryotic promoters contain Shine-Dalgarno sequences in addition to the -10 and -35 consensus sequences.

DNA "control sequences" refers collectively to promoter sequences, ribosome binding sites, polyadenylation signals, transcription termination sequences, upstream regulatory domains, enhancers, and the like, which collectively provide for the expression (i.e., the transcription and translation) of a coding sequence in a host cell.

A control sequence "directs the expression" of a coding sequence in a cell when RNA polymerase will bind the promoter sequence and transcribe the coding sequence into mRNA, which is then translated into the polypeptide encoded by the coding sequence.

A "host cell" is a cell which has been transformed or transfected, or is capable of transformation or transfection by an exogenous DNA sequence.

A cell has been "transformed" by exogenous DNA when such exogenous DNA has been introduced inside the cell membrane. Exogenous DNA may or may not be integrated (covalently linked) into chromosomal DNA making up the genome of the cell. In prokaryotes and yeasts, for example, the exogenous DNA may be maintained on an episomal element, such as a plasmid. With respect to eukaryotic cells, a stably transformed or transfected cell is one in which the exogenous DNA has become integrated into the chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones comprised of a population of daughter cell containing the exogenous DNA.

A "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth *in vitro* for many generations.

A "heterologous" region of a DNA construct is an identifiable segment of DNA within or attached to another DNA molecule that is not found in association with the other molecule in nature.

In accordance with yet a further aspect of the present invention, there is provided the use of a polypeptide of the invention for therapeutic or prophylactic purposes, for example, as an antibacterial agent or a vaccine.

In accordance with another aspect of the present invention, there is provided the use of a polynucleotide of the invention for therapeutic or prophylactic purposes, in particular genetic immunisation.

In accordance with yet another aspect of the present invention, there are provided inhibitors to such polypeptides, useful as antibacterial agents. In particular, there are provided antibodies against such polypeptides.

Another aspect of the invention is a pharmaceutical composition comprising the above polypeptide, polynucleotide or inhibitor of the invention and a pharmaceutically acceptable carrier.

In a particular aspect the invention provides the use of an inhibitor of the invention as an antibacterial agent.

The invention further relates to the manufacture of a medicament for such uses.

The polypeptide may be used as an antigen for vaccination of a host to produce specific antibodies which have anti-bacterial action.

The polypeptides or cells expressing them can be used as an immunogen to produce antibodies thereto. These antibodies can be, for example, polyclonal or monoclonal antibodies. The term antibodies also includes chimeric, single chain, and humanized antibodies, as well as Fab fragments, or the product of an Fab expression library. Various procedures known in the art may be used for the production of such antibodies and fragments.

Antibodies generated against the polypeptides of the present invention can be obtained by direct injection of the polypeptides into an animal or by administering the polypeptides to an animal, preferably a nonhuman. The antibody so obtained will then bind the polypeptides itself. In this manner, even a sequence encoding only a fragment of the polypeptides can be used to generate antibodies binding the whole native polypeptides. Such antibodies can then be used to isolate the polypeptide from tissue expressing that polypeptide.

Polypeptide derivatives include antigenically or immunologically equivalent derivatives which form a particular aspect of this invention.

The term 'antigenically equivalent derivative' as used herein encompasses a polypeptide or its equivalent which will be specifically recognised by certain antibodies which, when raised to the protein or polypeptide according to the present invention, interfere with the interaction between pathogen and mammalian host.

The term 'immunologically equivalent derivative' as used herein encompasses a peptide or its equivalent which when used in a suitable formulation to raise antibodies in a vertebrate, the antibodies act to interfere with the interaction between pathogen and mammalian host.

In particular derivatives which are slightly longer or slightly shorter than the native protein or polypeptide fragment of the present invention may be used. In addition, polypeptides in which one or more of the amino acid residues are modified may be used. Such peptides may, for example, be prepared by substitution, addition, or rearrangement of amino acids or by chemical modification thereof. All such substitutions and modifications are generally well known to those skilled in the art of peptide chemistry.

The polypeptide, such as an antigenically or immunologically equivalent derivative or a fusion protein thereof is used as an antigen to immunize a mouse or other animal such as a rat or chicken. The fusion protein may provide stability to the polypeptide. The antigen may be associated, for example by conjugation , with an immunogenic carrier protein for example bovine serum albumin (BSA) or keyhole limpet haemocyanin (KLH). Alternatively a multiple antigenic peptide comprising multiple copies of the the protein or polypeptide, or an antigenically or immunologically equivalent polypeptide thereof may be sufficiently antigenic to improve immunogenicity so as to obviate the use of a carrier.

For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler and Milstein, 1975, Nature, 256:495-497), the trioma technique, the human B-cell hybridoma technique (Kozbor et al., 1983, Immunology Today 4:72), and the EBV-hybridoma technique to produce human monoclonal antibodies (Cole, et al., 1985, in Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96).

Techniques described for the production of single chain antibodies (U.S. Patent 4,946,778) can be adapted to produce single chain antibodies to immunogenic polypeptide products of this invention.

Using the procedure of Kohler and Milstein (*supra* (1975), antibody-containing cells from the immunised mammal are fused with myeloma cells to create hybridoma cells secreting monoclonal antibodies.

The hybridomas are screened to select a cell line with high binding affinity and favorable cross reaction with other staphylococcal species using one or more of the original

polypeptide and/or the fusion protein. The selected cell line is cultured to obtain the desired Mab.

Hybridoma cell lines secreting the monoclonal antibody are another aspect of this invention.

Alternatively phage display technology could be utilised to select antibody genes with binding activities towards the polypeptide either from repertoires of PCR amplified v-genes of lymphocytes from humans screened for possessing anti-Fbp or from naive libraries (McCafferty, J. *et al.*, (1990), Nature **348**, 552-554; Marks, J. *et al.*, (1992) Biotechnology 10, 779-783). The affinity of these antibodies can also be improved by chain shuffling (Clackson, T. et al., (1991) Nature **352**, 624-628).

The antibody should be screened again for high affinity to the polypeptide and/or fusion protein.

As mentioned above, a fragment of the final antibody may be prepared.

The antibody may be either intact antibody of M_r approx 150,000 or a derivative of it, for example a Fab fragment or a Fv fragment as described in Skerra, A and Pluckthun, A (1988) Science **240** 1038-1040. If two antigen binding domains are present each domain may be directed against a different epitope - termed 'bispecific' antibodies.

The antibody of the invention may be prepared by conventional means for example by established monoclonal antibody technology (Kohler, G. and Milstein, C. *supra* (1975)) or using recombinant means e.g. combinatorial libraries, for example as described in Huse, W.D. *et al.*, (1989) Science 246,1275- 1281.

Preferably the antibody is prepared by expression of a DNA polymer encoding said antibody in an appropriate expression system such as described above for the expression of polypeptides of the invention. The choice of vector for the expression system will be determined in part by the host, which may be a prokaryotic cell, such as *E. coli* (preferably strain B) or *Streptomyces sp.* or a eukaryotic cell, such as a mouse C127, mouse myeloma, human HeLa, Chinese hamster ovary, filamentous or unicellular fungi or insect cell. The host may also be a transgenic animal or a transgenic plant [for example as described in Hiatt,A *et al.*,(1989) Nature 34, 76-78]. Suitable vectors include plasmids, bacteriophages, cosmids and recombinant viruses, derived from, for example, baculoviruses and vaccinia.

The Fab fragment may also be prepared from its parent monoclonal antibody by enzyme treatment, for example using papain to cleave the Fab portion from the Fc portion.

Preferably the antibody or derivative thereof is modified to make it less immunogenic in the patient. For example, if the patient is human the antibody may most preferably be 'humanised'; where the complimentarity determining region(s) of the hybridoma-derived antibody has been transplanted into a human monoclonal antibody , for example as described in Jones, P. *et al* (1986), Nature **321**, 522-525 or Tempest *et al*.,(1991) Biotechnology **9**, 266-273.

The modification need not be restricted to one of 'humanisation' ; other primate sequences (for example Newman, R. *et al*. 1992, Biotechnology,**10**, 1455-1460) may also be used.

The humanised monoclonal antibody, or its fragment having binding activity, form a particular aspect of this invention.

This invention provides a method of screening drugs to identify those which interfere with the proteins herein, which method comprises measuring the interference of the protein activity by test drug. For example, if the protein has enzymatic activity, after suitable purification and formulation the activity of the enzyme can be followed by its ability to convert its natural substrates. By incorporating different chemically synthesised test compounds or natural products into such an assay of enzymatic activity one is able to detect those additives which compete with the natural substrate or otherwise inhibit enzymatic activity.

The invention also relates to inhibitors identified thereby.

The use of a polynucleotide of the invention in genetic immunisation will preferably employ a suitable delivery method such as direct injection of plasmid DNA into muscles (Wolff *et al*., Hum Mol Genet 1992, 1:363, Manthorpe *et al*., Hum. Gene Ther. 1963:4, 419), delivery of DNA complexed with specific protein carriers (Wu *et al*., J Biol Chem 1989:264,16985), coprecipitation of DNA with calcium phosphate (Benvenisty & Reshef, PNAS,1986:83,9551), encapsulation of DNA in various forms of liposomes (Kaneda *et al*., Science 1989:243,375), particle bombardment (Tang *et al*., Nature 1992, 356:152, Eisenbraun *et al*., DNA Cell Biol 1993, 12:791) and *in vivo* infection using cloned retroviral vectors (Seeger et al, PNAS 1984:81,5849). Suitable promoters for muscle transfection include CMV, RSV, SRa, actin, MCK, alpha globin, adenovirus and dihydrofolate reductase.

In therapy or as a prophylactic, the active agent i.e the polypeptide, polynucleotide or inhibitor of the invention, may be administered to a patient as an injectable composition, for example as a sterile aqueous dispersion, preferably isotonic.

Alternatively the composition may be formulated for topical application for example in the form of ointments, creams, lotions, eye ointments, eye drops, ear drops, mouthwash, impregnated dressings and sutures and aerosols, and may contain appropriate conventional additives, including, for example, preservatives, solvents to assist drug penetration, and emollients in ointments and creams. Such topical formulations may also contain compatible conventional carriers, for example cream or ointment bases, and ethanol or oleyl alcohol for lotions. Such carriers may constitute from about 1% to about 98% by weight of the formulation; more usually they will constitute up to about 80% by weight of the formulation.

For administration to human patients, it is expected that the daily dosage level of the active agent will be from 0.01 to 10 mg/kg, typically around 1 mg/kg. The physician in any event will determine the actual dosage which will be most suitable for an individual patient and will vary with the age, weight and response of the particular patient. The above dosages are exemplary of the average case. There can, of course, be individual instances where higher or lower dosage ranges are merited, and such are within the scope of this invention.

A vaccine composition is conveniently in injectable form. Conventional adjuvants may be employed to enhance the immune response.

A suitable unit dose for vaccination is 0.5-5ug/kg of antigen, and such dose is preferably administered 1-3 times and with an interval of 1-3 weeks.

Within the indicated dosage range, no adverse toxicologicals effects are expected with the compounds of the invention which would preclude their administration to suitable patients.

EXAMPLES

In order to facilitate understanding of the following examples certain frequently occurring methods and/or terms will be described.

"Plasmids" are designated by a lower case p preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in

accord with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.

"Digestion" of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinarily skilled artisan. For analytical purposes, typically 1 µg of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 µl of buffer solution. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 µg of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37 C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the reaction is electrophoresed directly on a polyacrylamide gel to isolate the desired fragment.

Size separation of the cleaved fragments is performed using 8 percent polyacrylamide gel described by Goeddel, D. *et al.*, (1980) Nucleic Acids Res., 8:4057.

"Oligonucleotides" refers to either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides have no 5' phosphate and thus will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide will ligate to a fragment that has not been dephosphorylated.

"Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments (Maniatis, T., et al., Id., p. 146). Unless otherwise provided, ligation may be accomplished using known buffers and conditions with 10 units to T4 DNA ligase ("ligase") per 0.5 µg of approximately equimolar amounts of the DNA fragments to be ligated.

Example 1

Isolation of DNA from *S. Aureus* WCUH 29

The polynucleotide having the DNA sequence given in SEQ ID NO 1 was obtained from a library of clones of chromosomal DNA of *S.aureus* WCUH 29 in *E.coli*. In some cases the sequencing data from two or more clones containing overlapping *S.aureus* WCUH 29 DNA was used to construct the contiguous DNA sequence in Sequences set forth in SEQUENCE 1 [SEQ ID Nos:1,4,7,10,13,16,19,22,25,28,31,34,37,40,43,46,

49,52,55,58,61,64,67,70,73,76] of Table 1. Libraries may be prepared by routine methods, for example:

Methods 1 and 2

Total cellular DNA is isolated from *Staphylococcus aureus* strain WCUH29 (NCIMB 40771) according to standard procedures and size-fractionated by either of two methods.

Method 1.

Total cellular DNA is mechanically sheared by passage through a needle in order to size-fractionate according to standard procedures. DNA fragments of up to 11kbp in size are rendered blunt by treatment with exonuclease and DNA polymerase, and EcoRI linkers added. Fragments are ligated into the vector Lambda ZapII that has been cut with EcoRI, the library packaged by standard procedures and *E.coli* infected with the packaged library. The library is amplified by standard procedures.

Method 2.

Total cellular DNA is partially hydrolysed with a combination of four restriction enzymes (RsaI, PalI, AluI and Bsh1235I) and size-fractionated according to standard procedures. EcoRI linkers are ligated to the DNA and the fragments then ligated into the vector Lambda ZapII that have been cut with EcoRI, the library packaged by standard procedures, and *E.coli* infected with the packaged library. The library is amplified by standard procedures.

Example 2

The determination of expression during infection of a gene from *Staphylococcus aureus* WCUH29

Necrotic fatty tissue from a four day groin infection of *Staphylococcus aureus* WCUH29 in the mouse is efficiently disrupted and processed in the presence of chaotropic agents and RNAase inhibitor to provide a mixture of animal and bacterial RNA. The optimal conditions for disruption and processing to give stable preparations and high yields of bacterial RNA are followed by the use of hybridisation to a radiolabelled oligonucleotide specific to *Staphylococcus aureus* 16S RNA on Northern blots. The RNAase free, DNAase free, DNA and protein free preparations of RNA obtained are suitable for Reverse Transcription PCR (RT-PCR) using unique primer pairs designed from the sequence of each gene of *Staphylococcus aureus* WCUH29.

a) Isolation of tissue infected with Staphylococcus aureus WCUH29 from a mouse animal model of infection

10 ml. volumes of sterile nutrient broth (No.2 Oxoid) are seeded with isolated, individual colonies of Staphylococcus aureus WCUH29 from an agar culture plate. The cultures are incubated aerobically (static culture) at 37 degrees C for 16-20 hours . 4 week old mice (female,18g-22g, strain MF1) are each infected by subcutaneous injection of 0.5ml. of this broth culture of Staphylococcus aureus WCUH29 (diluted in broth to approximately 10^8 cfu/ml.) into the anterior , right lower quadrant (groin area). Mice should be monitored regularly during the first 24 hours after infection, then daily until termination of study. Animals with signs of systemic infection, i.e. lethargy, ruffled appearance, isolation from group, should be monitored closely and if signs progress to moribundancy, the animal should be culled immediately.

Visible external signs of lesion development will be seen 24-48h after infection. Examination of the abdomen of the animal will show the raised outline of the abscess beneath the skin. The localised lesion should remain in the right lower quadrant, but may occasionally spread to the left lower quadrant, and superiorly to the thorax. On occasions, the abscess may rupture through the overlying skin layers. In such cases the affected animal should be culled immediately and the tissues sampled if possible. Failure to cull the animal may result in the necrotic skin tissue overlying the abscess being sloughed off, exposing the abdominal muscle wall.

Approximately 96h after infection, animals are killed using carbon dioxide asphyxiation. To minimise delay between death and tissue processing /storage, mice should be killed individually rather than in groups.The dead animal is placed onto its back and the fur swabbed liberally with 70% alcohol. An initial incision using scissors is made through the skin of the abdominal left lower quadrant, travelling superiorly up to, then across the thorax. The incision is completed by cutting inferiorly to the abdominal lower right quadrant. Care should be taken not to penetrate the abdominal wall. Holding the skin flap with forceps, the skin is gently pulled way from the abdomen. The exposed abscess, which covers the peritoneal wall but generally does not penetrate the muscle sheet completely, is excised, taking care not to puncture the viscera

The abscess/muscle sheet and other infected tissue may require cutting in sections, prior to flash-freezing in liquid nitrogen, thereby allowing easier storage in plastic collecting vials.

b) Isolation of *Staphylococcus aureus* WCUH29 RNA from infected tissue samples

4-6 infected tissue samples(each approx 0.5-0.7g) in 2ml screw-cap tubes are removed from -80°C.storage into a dry ice ethanol bath In a microbiological safety cabinet the samples are disrupted individually whilst the remaining samples are kept cold in the dry ice ethanol bath. To disrupt the bacteria within the tissue sample 1ml of TRIzol Reagent (Gibco BRL, Life Technologies) is added followed by enough 0.1mm zirconia/silica beads to almost fill the tube, the lid is replaced taking care not to get any beads into the screw thread so as to ensure a good seal and eliminate aerosol generation. The sample is then homogenised in a Mini-BeadBeater Type BX-4 (Biospec Products). Necrotic fatty tissue is treated for 100 seconds at 5000 rpm in order to achieve bacterial lysis. *In vivo* grown bacteria require longer treatment than *in vitro* grown *S.aureus* WCUH29 which are disrupted by a 30 second bead-beat.

After bead-beating the tubes are chilled on ice before opening in a fume-hood as heat generated during disruption may degrade the TRIzol and release cyanide.

200 microlitres of chloroform is then added and the tubes shaken by hand for 15 seconds to ensure complete mixing. After 2-3 minutes at room temperature the tubes are spun down at 12,000 x g, 4 °C for 15minutes and RNA extraction is then continued according to the method given by the manufacturers of TRIzol Reagent i.e.:- The aqueous phase, approx 0.6 ml, is transferred to a sterile eppendorf tube and 0.5 ml of isopropanol is added. After 10 minutes at room temperature the samples are spun at 12,000 x g, 4 °C for 10 minutes. The supernatant is removed and discarded then the RNA pellet is washed with 1 ml 75% ethanol. A brief vortex is used to mix the sample before centrifuging at 7,500 x g, 4 °C for 5 minutes. The ethanol is removed and the RNA pellet dried under vacuum for no more than 5 minutes. Samples are then resuspended by repeated pipetting in 100 microlitres of DEPC treated water, followed by 5-10 minutes at 55 °C. Finally, after at least 1 minute on ice, 200 units of Rnasin (Promega) is added.

RNA preparations are stored at -80 °C for up to one month. For longer term storage the RNA precipitate can be stored at the wash stage of the protocol in 75% ethanol for at least one year at -20 °C.

Quality of the RNA isolated is assessed by running samples on 1% agarose gels. 1 x TBE gels stained with ethidium bromide are used to visualise total RNA yields. To demonstrate the isolation of bacterial RNA from the infected tissue 1 x MOPS, 2.2M formaldehyde gels are run and vacuum blotted to Hybond-N (Amersham). The blot is then

hybridised with a ^{32}P labelled oligonucletide probe specific to 16s rRNA of *S.aureus* (K.Greisen, M. Loeffelholz, A. Purohit and D. Leong. J.Clin. (1994) Microbiol. 32 335-351). An oligonucleotide of the sequence:-

5'-gctcctaaaaggttactccaccggc-3' [SEQ ID NO:91]

is used as a probe. The size of the hybridising band is compared to that of control RNA isolated from *in vitro* grown *S.aureus* WCUH29 in the Northern blot. Correct sized bacterial 16s rRNA bands can be detected in total RNA samples which show extensive degradation of the mammalian RNA when visualised on TBE gels.

c) The removal of DNA from Staphylococcus aureus WCUH29 derived RNA

DNA was removed from 73 microlitre samples of RNA by a 15 minute treatment on ice with 3 units of DNAasel, amplification grade (Gibco BRL, Life Technologies) in the buffer supplied with the addition of 200 units of Rnasin (Promega) in a final volume of 90 microlitres.

The DNAase was inactivated and removed by treatment with TRIzol LS Reagent (Gibco BRL, Life Technologies) according to the manufacturers protocol.
DNAase treated RNA was resuspended in 73 microlitres of DEPC treated water with the addition of Rnasin as described in Method 1.

d) The preparation of cDNA from RNA samples derived from infected tissue

10 microlitre samples of DNAase treated RNA are reverse transcribed using.a SuperScript Preamplification System for First Strand cDNA Synthesis kit (Gibco BRL, Life Technologies) according to the manufacturers instructions. 1 nanogram of random hexamers is used to prime each reaction. Controls without the addition of SuperScriptII reverse transcriptase are also run. Both +/-RT samples are treated with RNaseH before proceeding to the PCR reaction

e) The use of PCR to determine the presence of a bacterial cDNA species

PCR reactions are set up on ice in 0.2ml tubes by adding the following components:

45 microlitres PCR SUPERMIX (Gibco BRL, Life Technologies).

1 microlitre 50mM $MgCl_2$, to adjust final concentration to 2.5mM.

1 microlitre PCR primers(optimally 18-25 basepairs in length and designed to possess similar annealing temperatures), each primer at 10mM initial concentration.

2 microlitres cDNA.

PCR reactions are run on a Perkin Elmer GeneAmp PCR System 9600 as follows: 5 minutes at 95 °C, then 50 cycles of 30 seconds each at 94 °C, 42 °C and 72 °C followed by 3 minutes at 72 °C and then a hold temperature of 4 °C. (the number of cycles is optimally 30-50 to determine the appearance or lack of a PCR product and optimally 8-30 cycles if an estimation of the starting quantity of cDNA from the RT reaction is to be made).

10 microlitre aliquots are then run out on 1% 1 x TBE gels stained with ethidium bromide with PCR product, if present, sizes estimated by comparison to a 100 bp DNA Ladder (Gibco BRL, Life Technologies). Alternatively if the PCR products are conveniently labelled by the use of a labelled PCR primer (e.g. labelled at the 5'end with a dye) a suitable aliquot of the PCR product is run out on a polyacrylamide sequencing gel and its presence and quantity detected using a suitable gel scanning system (e.g. ABI Prism™ 377 Sequencer using GeneScan™ software as supplied by Perkin Elmer)

RT/PCR controls may include +/- reverse transcriptase reactions, 16s rRNA primers or DNA specific primer pairs designed to produce PCR products from non-transcribed *S.aureus* WCUH29 genomic sequences.

To test the efficiency of the primer pairs they are used in DNA PCR with WCUH29 total DNA. PCR reactions are set up and run as described above using approx. 1 microgram of DNA in place of the cDNA and 35 cycles of PCR.

Primer pairs which fail to give the predicted sized product in either DNA PCR or RT/PCR are PCR failures and as such are uninformative. Of those which give the correct size product with DNA PCR two classes are distinguished in RT/PCR:

1.Genes which are not transcribed *in vivo* reproducibly fail to give a product in RT/PCR.

2.Genes which are transcribed *in vivo* reproducibly give the correct size product in RT/PCR and show a stronger signal in the +RT samples than the signal (if at all present) in -RT controls.

The following nucleotide sequences (sequences set forth in SEQUENCE 1 [SEQ ID Nos:1,4,7,10,13,16,19,22,25,28,31,34,37,40,43,46,49,52,55,58,61,64,67,70,73,76] of Table 1) were identified in the above test as transcribed *in vivo*. Each set of sequences relates to a separate gene (Gene #). Deduced amino acid sequences are given where available as the sequences set forth in each SEQUENCE 2 [SEQ ID Nos: 79,80,81,82,83,84,85,86,87,88,89,90] of Table 1. The pair of PCR primers used to

identify the gene are given as the sequences set forth in SEQUENCE 3 [SEQ ID Nos: 2,5,8,11,14, 17,20,23,26,29,32,35,38,41,44,47,50,53,56,59,62,65,68,71,74,77] and 4 [SEQ ID Nos: 3,6,9,12,15,18,21,24,27,30,33,36,39,42,45,48,51,54,57,60.63,66,69,72,75,78] of Table 1. Homologies to known genes are given where determined and represent the putative identification of gene function for each gene in Table 1.

TABLE 1

Gene #1
E.coli pts system 5'end ptfB

SEQUENCE 1 [SEQ ID NO:1]

```
       1  CTAGGAGTAG TATTTGGTTC ATGATTGCCT AATTCAATCA CATCTTTACT
      51  TTGCTCTAAG TGCAAATCAC GCAATTGACC ATNTGGATCT CGTCTATCAT
     101  AGTCATAAAT ACGGTATGTC GTATCGGATG ATTGTTGTGT CTCTAAAATT
     151  AAAATACCCG AACCAATGGC ATGGACAGTG CCAGCAGGAA CATAATAAAA
     201  GTCACCGGGC TTAACAGGTA TACGTTTGAA AAGACTGCCA AATTCATGAT
     251  TATCAATCAT GTCGATTAAC GCCTGTTTAT TATGTGCATG GACGCCATAA
     301  TATAATTTCA GCACCTGGGC TGCATCTAAA TATACCAACA TTCTGTTTTA
     351  CCTAGTTCGC CTTCGTGTTT TAAAGCGTAG TCATCATCTG GATGAACTTG
     401  AACAGATAAT TTATCATTGG CATCTAATAC TTTAGTTAGC AGAGGGAAAC
     451  TATCTCGTGA ATCATTATCG AATAATTCAC GATGTTGTGA CCAAAGTTGA
     501  TCTAGGGTCA TATCCTTGTA TGGACCATTG ATAATTGTAT TAGGACCATT
     551  TGGATGTGCA GAAATTGCCC AGCATTCACC AGTTGTTTCA TTAGGGATAT
     601  CATAGTTAAA TGCTTTTAAT GCATGACCGC CCCAAATTCT GTCTTTAAAA
     651  ACGGGTTGTA AAAATAATGC CATAGTTAAA ACTCCTCTAT ATTTTCATTA
     701  ATAAGTTATA AATTTCTGTA GTACTGTTGG CATTAATTAG TGATTGGCGT
     751  GTCTCATCAT TCATTAACGC TTTAGATAAG CGCTGAAGTA TTTTTAAATG
     801  TGTATCCTGA CTGTTGTTTG GTACGGCAAT TAAGAATATC AATTGAGGTA
     851  GACTACCATC TAGACTGTCC CATTTAACAC CATGATTATT TTTCATAACA
     901  GCTACAATCG GTTGTTTTAC AACATCAGAC TTTGCATGTG GAATGGCCAC
     951  GTTCATGCCA ATAGCTGTCG TAGACTCCAT TTCACGTTCT AGTATTGCAT
```

1001 TTTTTAAATG CGATGTGTGC TCTACATAAC GGCAAATTTT AAGTTTATGA
1051 ATCAACATAT CAATTGCTTC GTTTCGAGAC ATGTCGTGAT CAGTAATTAT
1101 CATAGTTTGT TGATCAAAAA CATGAGAAGG TTTATTGAGA TGTGAATGTT
1151 TCGCTCGTGC CATCNACATT GTCAACCTCT GTATCATGTT GTGTAATATC
1201 TGTATCATGA AGTTGCGTGT GTTGCGCTGG TGCATCTACT GCTATAACTG
1251 GTGTATTGCG TNTTAATAAT AGTACAGTAG GCATTGTGAC AAGACTACCT
1301 ACTATCNCTC CAAAGATAAA CCATAATACA TGATCAATAC CACCTAATAC
1351 AGCCACGATT GGACCTCCAT GTGCGACTCT ATCGCCGACA CCACCAATGN
1401 CTGCAATGAC TGATGCAATC ATTGCACCAA TGATGTTTGC AGGTATAATG
1451 CGCAATGGAT CTTGGGCTGC GAAAGGAATA GCACCTTCAG TAATNCCAAA
1501 TAGTCCCATA GTGAAGGNAG CCTTACCCAT TTCTCTTTCG GAATGATTGA
1551 ATTTATACTT NTGAACANAC GTTGCTAAAC CTAAACCGAT TGGTGGTGTA
1601 CATACANCAA CTGCGACCAT ACCCATAACG GCGTAATTAC CTTCAGCAAT
1651 AAGTGCTGAG CCAAATAAAA ATGCTACCTT GTTTAATTGG ACCGCCCATA
1701 TCGAAGGCGA TCATCGCACC TATAATCATC GACAAGTATA ATAATATTAG
1751 CACCTTGCAT ACTTTTTAAC CAGGGTTGTT AGGAATGCCG CAAAAATATT
1801 AGAAATCGTG CACCGATTAA AAATATAAAT ATCAATCCTA ACAACGACCG
1851 ATGAAATAAT GGGAATAATA ATGATAGGCA TAATTGGTGC CATTGCTTTT
1901 GGAACTTTAA TATCTTTAAT CCACTTTGCG ATATAACCTG CTAAGAAACC
1951 AGCAACAATA CCACCTAAAA ATCCTGCGCC TGCATCACTG CCATAAAAAC
2001 TACCGTCAGC AGCGATAGCG CCGCCAATCA TACCAGGAAC AAGACCGGGC
2051 TTGTCAGCGA TACTAACAGC GATATATCCA GCTCGTGCCG AATTCGGCAC
2101 GAGCTCGTGC C

SEQUENCE 2 (STOPS SHORT) [SEQ ID NO:79]
1 MGMVAVXVCT PPIGLGLATX VXKYKFNHSE REMGKAXFTM GLFGITEGAI
51 PFAAQDPLRI IPANIIGAMI ASVIAXIGGV GDRVAHGGPI VAVLGGIDHV
101 LWFIFGXIVG SLVTMPTVLL LXRNTPVIAV DAPAQHTQLH DTDITQHDTE
151 VDNVDGTSET FTSQ*

SEQUENCE 3 [SEQ ID NO:2]
accctctgta tcatgttg

SEQUENCE 4 [SEQ ID NO:3]
gtgcgatgat cgccttgg

Gene #2
E.coli RelA

SEQUENCE 1 [SEQ ID NO:4]

```
       1  CGGCTCTTCG TAATATTGAT AATGTGCAAT ATTTNAAGAA TAATCAATTT

      51  ATTGAAGAAG AAACCGTAGT GACCGTGAGC GAATATCGAA NCGGCTATTG

     101  ATAGAATACG TACTGAAATG GACCCGAATG AATATCGAAG NCGATATAAA

     151  TGGTAGACCT AAACATATTT ACAGTATTTA TCGGNAAATG ATGAAGCAGA

     201  AAAAACAATT TGATCAAATT TTTGATTTGT TGGCGATACG TGTTATTGTC

     251  AATTCTATTA ATGATTGTTA TGCGATACTT GGGTTGGTGC ATACGTTATG

     301  GAAACCGATG CCAGGACGTT TTAAAGATTA TATTGCAATG CCTAAACAAA

     351  ATTTGTATCA GTCATTGCAT ACTACAGTAG TAGGTCCAAA TGGAGACCCG

     401  CTCGAAATCC AAATACGAAC GTTTGATATG CACGAAATTG CTGAGCATGG

     451  TGTTGCAGCA CACTGGGCTT ACAAAGAAGG TAAAAAAGTA AGTGAAAAAG

     501  ATCAAACTTA TCAAAATAAG TTAAATTGGT TAAAAGAATT AGCTGAAGCG

     551  GATCATACAT CGTCTGACGC TCAAGAATTT ATGGAAACCT TATAATATGA

     601  CTTACAGAGT GACAAAGTAT ACGCATTTAC CCCAGGGAGT GATGTTATTG

     651  AGTNGGCATA TGGTGCTGTG CCGATTGGAT TTTGGCTTAT GCGAATCACA

     701  GGGAANGTAG GTAATAAGAT GATTGGCGCC CAGGTGGAAT GGCAAAATTG

     751  TACCANATTG ACTTATNTTT TCACAAAACA GGCGGATATT GTTGGAAATA

     801  CCGTTCTAG
```

SEQUENCE 2 [SEQ ID NO:80]

```
       1  MNIEXDINGR PKHIYSIYRX MMKQKKQFDQ IFDLLAIRVI VNSINDCYAI

      51  LGLVHTLWKP MPGRFKDYIA MPKQNLYQSL HTTVVGPNGD PLEIQIRTFD

     101  MHEIAEHGVA AHWAYKEGKK VSEKDQTYQN KLNWLKELAE ADHTSSDAQE

     151  FMETL*
```

SEQUENCE 3 [SEQ ID NO:5]
agatacgtac tgaaatgg

SEQUENCE 4 [SEQ ID NO:6]
cctgtgattc gcataagc

Gene #3
Staph FemB

SEQUENCE 1 [SEQ ID NO:7]

```
   1  GTGATGTGGC TAAACGCTTA AATGCAAATA TATATGTGTC TGGCGAAGGT
  51  GAAGATGCAT TAGGGTATAA AAATATGCCA TCAAAAACAC AATTTGTTAA
 101  ACATGGAGAT ATCATTCAAG TAGGCAATGT TAAATTAGAA GTTCTGCATA
 151  CTCCAGGACA CACGCCTGAA AGTATTAGCT TTTTACTCAC TGATTTAGGT
 201  GGTGGNTCAN GTGTTCCGAT GGGATTATTT AGTGGTGACT TTATTTNTGN
 251  TGGTGATATA GGTAGACCTG ATTTATTAGA AAAATCTTGT TCAAATAAAG
 301  GGTTCGGCAC GAAATTAGCG CGAAACAAAT GTATGAGTCC GATCAAAATA
 351  TTAAAAATTT ACCAGACTAT GTTCAAATCT GGCCGGGTCA TGGTGCTGGA
 401  AGCCCTTGTG GTAAAGCATT AGGTGCCATA CCTATATCTA CAATAGGTTA
 451  TGAGAAAATT AATAACTGGG CATTTAATGA AATTGATGAG ACTAAATTTA
 501  TTGNNTCATT AACATCAAAT CAACCAGCAC CACCNCATCA TTGTGCACAA
 551  ATGAAACAAG TTANTCAGTG TGGCATGAAT TTATNTCAAT CATATGATGT
 601  TTATCCNAGC TTAGATNATA AGAGAGTAGC ATTTGATCTT CGCGTAGCAA
 651  AGAGGGCTTT CACGGGTGGC CACACAAAAG GAACAATCAA TATACCATAC
 701  AACAAAAACT TTATTANTCA ANTTGGGTGG GTACTTAGAT TNTGAAAAAG
 751  ATATAGATTT AATTGGAGAT AAATCTACTG TTGAGAAAAG CGAAACACAC
 801  TTTACAATTA ATTGGGTTTG ATAAGGTAGC AGGCTATCGT NTGCCAAAAT
 851  CAGGCATTTC ACCCCAGTCC GNTCATAGCG CTGATATGAC AGGTAAAGAA
 901  GAACATGTAT TAGACGTACG TAATGATGAA GAGTGGAATA ATGGACACTT
 951  AGNTCAAGCA GTTAATATTC CACATGGTAA ATTATTAAAT GAAAATATTC
1001  CTTTTAATAA AGAGGATAAA ATATATGTAC ATTGTCAGTC AGGTGTTAGA
1051  AGNTCAATTG CAGTGGGGTA TATTGGGAAA GCAAAGGCTT
```

SEQUENCE 2 [SEQ ID NO:81]

```
       1  DVAKRLNANI YVSGEGEDAL GYKNMPSKTQ FVKHGDIIQV GNVKLEVLHT

      51  PGHTPESISF LLTDLGGGSX VPMGLFSGDF IXXGDIGRPD LLEKSCSNKG

     101  FGTKLARNKC MSPIKILKIY QTMFKSGRVM VLEALVVKH*
```

SEQUENCE 3 [SEQ ID NO:8]

ttcgggtgtt ttaccttc

SEQUENCE 4 [SEQ ID NO:9]

tgcagcaagc cttttctc

Gene #4

DiCitrate Binding Protein

SEQUENCE 1 [SEQ ID NO:10]

```
       1  AGCAGAATCT TTTTTAGCAT GATCTGTCAT AATGATCATA CGCTCTGGAT

      51  TTAAATCAGC TAAATGTTCA GTGTCTAATT GTAAGTAAGG TCCTTTCAAA

     101  TATTTACTTA AACCTTGTGT TACATCGTCA CTTAATGCAT TTTTAAATCC

     151  TAGNTCGTTT AAAAATTGTC CAACATATGA ATAGTGTGGA TGTGCTAATA

     201  AACCAGCTTT AGCAACTACT GCTGGAAGCA CTTTGTGATT TCTATCAAAT

     251  TTAATTTCAT CTTTATACTT ATTGATTAAT TTATCATGCT CAGCAAGACG

     301  TTTNNCGCCT TCTTTNTCTT TATTTAAAGC TTTAGCAATT GTTGTTGAAC

     351  GAATTAATAT TGTGGGTGTA GTCTCCATCA AAACTCTTTA ATGATAATGT

     401  GGTGCAATGT GGGCTAATTC TTTATTAATA CCCTTATGTC TACTGCTATC

     451  AGNGATAATT AATCCCGGNT TTAATTTACT AATNTCTCTT AAGTTNGCTT

     501  GTTACGTGTA CCTACAGAAG TATTACCCCC AATTTTTCTC TTACTGGGTT

     551  ATGATACGTT TTTTCTTACC ATCATCAGCA ATACCAACTT GGTNTAACGG

     601  CTATATGCTG NTAATGCAAC CTTGCAAATG AGTACTCTAA TACAACGATA

     651  CGTTGTGCAT CTTTAGGTAC TTTTACTGTA CCATTTTCAT CTTTTACCCG

     701  AAATAGTATC TTTAGTTGAT GATTCTTCTT TTACTTGAAT TATCCGTATT

     751  ACCACAAGCT GCAACTAAAA GTAAGGCAAC TATTAATCCC AATATACTAA

     801  AAGTTTTTAG ACCTCTCATC NGTCCCACTC CTTAATATGT ATANCTTCAT

     851  TTATTATTTT ATTGATAACA ATTATCATTG TCAAGTAGCG TTCAATCTTT

     901  TTTATATTTC TAAAATGTAT GACTATATAT TTCCTCTAAT AATTATGACT
```

951 ACAATTAGCA CATTTCCTTA GACAAAATAC TGATAATGTA TCATTGCTAT

1001 ATCATCTTTG CATTAATACA ATTGACACCA CTTAGCATGA CCGNTATCCC

1051 TGTAATTCAG CTGATATTAT CTGTTGCAAT TTTATGTGAC GAACTGTTGC

1101 ACTTAATTTG ATAANTCAAC AANTACAANA NATCTAAGTT GAACAATTAT

1151 GATACAACCG TGCAAACGAT ATGTAGTATA ACTTGTCAAC TTAGAATTAT

1201 TGATAAATAT ATTAATATTG GTTTACCATA GCAGGAGATT TCACATCAAA

1251 ATTTTGAAGT AGCGTATCAA TCTTTGAATC ATCAATATAT ACCTTATGTA

1301 AATTTTTCAT ATACATCGAA TGAGAAAGTG CTTCATAATT TAATGAAAAA

1351 GATATATGAT CTCCAACTTG ATAGTGTCCT TGACCATTTA AATCAAGCAT

1401 TAAATGATCA CTCGAAGCGC CTAAAATATT GATATGCTGA TCCATAGGTG

1451 AAATATTATC GACTTGTGTA TCTNAAATAA CCAATATCTA CAATAGCTTG

1501 TAAGAATGAT TCATGCGTGT GTGTATTAAC TCGAGGTTTA ATTTCTAAAA

1551 TCTCAGCCTC CAATGTAATC GCATCTTGAT ATAACATAGC GAATCGCTTG

1601 ATTTGCGTTG TTTCAACAAC TCTAAACAAC GTNTCANCTA TTCGGAANTC

1651 AATTTATTTT TACCCAAATC AATATATAAA AGGTGGGGGG NAACATGCTC

1701 CGAATTACCA CCCGGAAATA ATTTNCANTC GATATCCTAT TTCTCTTNCA

1751 ACAGCTGAGA CGAATCGATT AATCATAAAG ATATCANCAC CACTTGGCGC

1801 ATCAGATTTA AAACACATAA AATTGAATGC TAAACCTACA AAATGGATAT

1851 TTTNCAAGTG AATAATCTCT TTANTATAAT CTAAAACATC ATAAGTCAGA

1901 ACACCTTCAC GGACATCTTT CCAATCTACC ATTAATAAAA TCTTATGTTT

1951 TTTTCCTAAA ACTTCTGCTA CTTCATTTAT NTGATGTATG GTAGATAATT

2001 CTGTGTGGAT ACTCATATCA ACTTTCCTCT ATCATATCTG AAATCTCTTT

2051 TGNGGGAGGC GTACGCAATA ACGTATATGT TAAATCCTGA TCTGCAATAC

2101 TAATTATGTT ATCCAATCTG GATTCTGCAA CATGATTGAT ACCTAACGCT

2151 TTTAAGCTTN CTACAATGGT ACGGGCANCA GCTATACACT TAATTACTGG

2201 TGTGANTNGN ATATTTTTAC TTTGAAAACT NNGTGGAGGT ACTTGGG

SEQUENCE 3 [SEQ ID NO:11]
tgtaagtaag gtcctttc

SEQUENCE 4 [SEQ ID NO:12]
taatacttct gtaggtac

Gene #5
Staph enterotoxin etxA

SEQUENCE 1 [SEQ ID NO:13]

```
   1  GGCACGAGCG GCACGAGCGT GTTGTATCAA GATTTTGTAG GCAGTTTTAC
  51  AACGTCCGAT TCAGCAAGTT ATGCACAAGA TTTTAAATCT GAGGAAAACG
 101  CTAAAAAGAT TGCTGAAACT TTAAATCTTT TATATCAATT AACAGGCAAT
 151  CAAAACGGTG TGAAAGTTGT GAAAGAAGTT GTGGATAGAA CTGACTTGTC
 201  ATCTGATAAA TCAGTTGATA GCGAAACAAT GTAACTATAC TAAGTTATGA
 251  GCATTACGCT CATAGCTTTC TTAGAAAGTA GGTGTAGTTT TGGATGATAT
 301  TCAGAAAATA AAAAAAGAGC TTTCTGAATT AGTTGAACGT GTTGATGATG
 351  TTGAAATACT AGCAAACGAA ACAGCTGATC ATGTGCTTGA ACTTAGAGAG
 401  GAACATAAGC AACATCATAA TGAACTAAGA GAATCTCATA AAGAACTTAA
 451  AGATAAGCAA GATAAAGTTG TAGATGAGAA TTTAGAGCAA ACAAAGATAT
 501  TAAACAGAAT TGAAGAAAGA TATCANACGC AAGTAGNTGT TGNGCAAAAA
 551  AATGAAGAAA AGACACTCGC CCAAAATAAA TGGCTCGTAG GTGCCATATG
 601  GGCGCTTGTA ACAATTGTTA TGATTGCAGT CATTACTGCA TCAATTNCTG
 651  CGTTATTACC TTAAGGGAGG TGGACATAAT GAGTTGGGCA AGATGGTTAT
 701  CATGTTATTT GTNTGGTCGT AAATGTAAAT AATGTTTTTG GTCAGTGCAT
 751  CGGCACTGGC TTTTTATTTT GATTGAAAAG AGGTACGTAC ATGGTATTAC
 801  ACAGCTCACA AGACAGGAAG CATACTCCAA GTGAAGTTGG GAAGTGTTGT
 851  TAATACCAAG TAAGTAGGAT ATCTGANATG TATAATAGAG TAAAAATGAA
 901  ATCTTTTTAT TATAGACACA TATAAAAAGT GTATAGTAAT ATATGTATGT
 951  ATAATTAAAT GATAATCATT TCATAATTAT TGTATATAAC TAAATAACTA
1001  CTTAACANAA ATAATTATGC TTTAGAGNTG ACCANNATGA NNNANNCCAG
1051  CATTTACATT ACTTTTATTC ATTGCCCTNA CGTTGACNAC AAGTCCCANT
1101  TGTAAATGGT AGCGAGAAAA GCGNAGNAAT AAATGCGAAA GATTTGCGAA
```

1151 AAAAGTCTGA ATTCCAGGGN ACAGCTTTAG NCAATCTTAN NCANATCTAT

1201 TATTACNATG NNANAGCTAN AACTGAAAAT AAAGAGAGTC CNCGACCACA

1251 TTTTTACAGC ATACTATATT GTTTANAGGC TTTTTTACAG ATCATTCGTG

1301 GTATANCGAT TTATTAGTAG ATTNTGATTC NNAGGATATT GTTNATAAAA

1351 ATAAAGGGNA AANAGTAGAC TTGTATGGTG CTTATTATGG TTATCAATGT

1401 GCGGGTGGTA CACCACACAA AACAGCTTGT ATGTATGGTG GTGTAACGTT

1451 ACATGATAAT AATCGATTGA CCGAAGAGAA AAAAGTGCCG ATCAATTTAT

1501 GGCTAGACGG TAAACANAAT ACAGTACCTT TGGAAACGGT TAAAACGAAT

1551 AAGAAAAATG TAACTGTTCA GGAGTTGGAT CTTCAAGCAA GACGTTATTT

1601 ACAGGAAAAA TATAATTTAT ATAACTCTGA TGTTTTTGAT GGGAAGGTTC

1651 AGAGGGGATT AATCGTGTTT CATACTTCTA CAGAACCTTC GGTTAATTAC

1701 GATTAATTTG GTGCTCAAGG ACAGTATTCA NATACACTAT TAAGAATNTA

1751 TAGAGATAAT AAAACGATTA ACTCTGAAAA CNTGCGTAG

SEQUENCE 2 (Short) [SEQ ID NO:82]
1 MYGGVTLHDN NRLTEEKKVP INLWLDGKXN TVPLETVKTN KKNVTVQELD

51 LQARRYLQEK YNLYNSDVFD GKVQRGLIVF HTSTEPSVNY D*

SEQUENCE 3 [SEQ ID NO:14]
atcccctctg aaccttcc

SEQUENCE 4 [SEQ ID NO:15]
aaatggtagc gagaaaag

Gene #6
Staph Lipase Precursor

SEQUENCE 1 [SEQ ID NO:16]
1 TCAAATGCAG TCAGGGAAGC AATAGGACGA TATGCATAAA GGAGATGGTA

51 AAGTGGAACA GTGACAGAAG GTAAAGACAC GCTTCAATCA TCGGAGNCAT

101 CAATCAANCA CAAAATAGTA AAACAATCAG GAACGCAAAA TGATAATCAA

151 GTAAAGCAAG ATTCTGGAAC GACAAGGTTC TAAACAGTCA CACCAAAATA

201 ATGCGACTAA TAATACTGAA CGTCAAAATG ATCAGGTTCA AAATACCCAT

251 CATGCTGAAC GTAATGGATC ACAATCGACA ACGTCACAAT CGAATGATGT

301 TGATAAATCA CAACCATCCA TTCCGGCACA AAAGGTATTA CCCAATCATG

```
 351  ATAAAGCAGC ACCAACTTCA ACTACACCCC CGTCTAATGA TAAAACTGCA
 401  CCTAAATCAA CAAAAGCACA AGATGCAACC ACGGACAAAC ATCCAAATCA
 451  ACAAGATACA CATCAACCCG CGTGCCTCAA ATCATAGATG CAAAGCAAGA
 501  TGATACTGTT CGCCAAAGTG AACAGAAACC ACAAGTTGGC GATTTAAGTA
 551  AACATATCGA TGGTCAAAAT TCCCCAGAGA AACCGACAGA TAAAAATACT
 601  GATAATAAAC AACTAATCAA AGATGCGCTT CAAGCGCCTA AAACACGTTC
 651  GACTACAAAT GCAGCAGCAG ATGCTAAAAA GGTTCGACCA CTTAAAGCGA
 701  ATCAAGTACA ACCACTTAAC AAATATCCAG TTGTTTTTGT ACATGGATTT
 751  TTAGGATTAG TAGGCGATAA TGCACCTGCT TTATATCCAA ATTATTGGGG
 801  TGGAAATAAA TTTAAAGTTA TCGAGGGAAT TGAGAAAGCA AGGCTATAAT
 851  GTACATCAAG CAAGTGTAAG TGCATTTGGT AGTAACTATG ATCGCGCTGT
 901  AGAACTTTAT TATTACATTA AAGGTGGTCA CGAGCGTAGA TTATGGCGCA
 951  GCACATGCAG CTAAATACGG ACATGAGCGC TATGGTAAGA CTTATAAAGG
1001  AATCATGCCT AATTGGGAAC CTGGTAAAAA GGTACATCTT GTAGGGCATA
1051  GTATGGGTGG TCAAACAATT CGTTTAATGG AAGAGTTTTT AAGAAATGGT
1101  AACAAAGAAG AAATTGCCTA TCATAAAGCG CATGGTGGAG AAATATCACC
1151  ATTATTCACT GGTGGTCATA ACAATATGGT TGCATCAATC ACAACATTAG
1201  CAACACCACA TAATGGTTCA CAAGCAGCTG ATAAGTTTGG AAATACAGAA
1251  GCTGTTAGAA AAATCATGTT CGCTTTAAAT CGATTTATGG GTAACAAGTA
1301  TTCCGAATAT CGATTTAGGA TTAACGCAAT GGGGCTTTAA ACAATTACCA
1351  AATGAGAGTT ACATTGACTA TATTAAAACG CGTTAGTAAA AGCAAAATTT
1401  GGACATCAGA CGATAATGCT GCCTATGATT TAACGTTAGA TGGCTCTGCA
1451  AAATTGAACA ACATGACAAG TATGAATCCT AATATTACGT ATACGACTTA
1501  TACAGGTGTG TCTTCACATA CTGGTCCATT AGGGCACGAA AATCCTGCCG
1551  AATTAGGCAC GAGACATTTT TCTTAATGGA TACAACGAGT AGAATTATTG
1601  GTCATGATGC AAGAGAAGAA TGGCGTAAAA ATGATGGTGT CGTACCAGTG
1651  ATTTCGTCGT TACATCCATC CAATCAACCA TTTATTAATG TTACGAATGA
```

1701 TGAACCTGCC ACACGCAGAG GTATCTGGCA AGTTAAACCA ATCATACAAG

1751 GATGGGATCA TGTCGATTTT ATCGGTGTGG ACTTCCTGGA TTTCAACACC

1801 GTAAGGTGCA GAACTTGCCA ACTTCTATAC AGGTATAATA AATGACTTGT

1851 TGCGTGTGGA AGCGNCTGAA AGTAAAGGAA CACAATTGAA AGCAAGTTAA

1901 ATTCATCTTC TGAATTTAAT AGGCTATGTA AATCGTGCTG TTATCATGGC

1951 ACATCAGATA TAAGTAGCAT CACAGTGTTG AATCTCAAAA TAGTAAAGTG

2001 AAATAAAGCG CCTGTCTCAT TAGCGAAAAC TAAAGGGACA GGCGTATCTG

2051 TTTATGAGCT TAATAAATTG TATGAATAAT ATGGTTGATC GAATAACTGT

2101 TTATCATTGA TGATAAATTT GAGTTTTTTA AAAATAATTG ATATATTACA

2151 CCATTGTTAT AGCGTTTAAA GAAATCAACC CAACTTTACG ATAAATAGTG

2201 ATTGCTTCGT CATTAGGTCT ACGATCAAAA TCATGCTCGT TTTTATTCAC

2251 GCGTTCAAAT GTTGAATGTG GAACATGATT CATGATATGT TCGCTTTCCT

2301 CAACGGGAAC ATCATAATCG CCATTACAAT GCGCAATGAA AACAGGTGGA

2351 AGTGTTTTAA GNTCATCTGG TGCAATATTA TATTTTGAAT CAGTATAATC

2401 ANCAATGTTA ATCATATTTA TCCATTTACC TGTGCCACGT GCATAAACGT

2451 AGAGTAAAAA ACGTGTGCGA TTTGATCTTG ANCAACCGGT GTTGGTGAAG

2501 TGAGTTGTCC AATCATTGTT TCGTTTATGC TTTGAGCTAT TTTTGCGTAA

2551 TACCTATTAG TTGTTTTAAA AGGGTTCAGT GTTGATGCGA CTATAACCAT

2601 AAAAATCAAT AACACCATCA ATATCTCTGT CTCGTGCAAT TAATAAGACT

2651 TAAATATGCA CCTGATGATC TGCCAAAGGT AAAAATAGGG CAATTAGAAT

2701 ATTGTGATTG AATCGCATCG AATGATGCGT AGACATCCTC AATAATGCAA

2751 TCGAGACTTA CTTCTGGTAA TAAACGATAA CTTAGTTGAA TTAAATCGTA

2801 ATGTTCCGTA AGGATATCGA TATACTGTGG GGATAAATCG TTAGCTTTAC

2851 CGAACATTAA TCCACCACCG TGGATGTAGA CAATAACGCC TTTTGTTGGT

2901 TGATTTTTTG CTTTAATAAT TGTGTAAGGT AATGCAAATG CATCTTTAGT

2951 AATTACTTTA TATTTAATTT CAGTCACGAT TTAATAGGCT CCTTAGGAAT

3001 CCGATATTGA TGTCATTATA ACACTGTCNT NAATTTCCAT GNAAAATAGT

3051 CTTAAGACGA TGAGTCATGA TAATTCTGTT CCAATTGACG TAAAGCGTCN

```
3101  CGGGTATGCT TCTTTAGACC TTCCCCATAA TCCATCATTT TAACAATATC
3151  TTTAAAAGCA GCATGTGGNA TGGCTAAATC TTCTAAATCT GCCATAGAAA
3201  ATTCAAGATT GATATCATGT GGTCGCTGTT CAGCAAGTTT ATGCACAAAG
3251  TCAGGTTCTG TGACCAAAGG CGAAGACATG CCGACCATAT CTGCATGTTG
3301  TAAAGCATCT AAAGCAGACT CTGGAGAATT AATCCCGCCA CTTGCAATTA
3351  AAGGGATACG ACCTGCTAAA TGTTCATAGA CAATTTGGTT AACTGGTCGA
3401  CCGAAATGAT CACCTGGTGT ACGAGACGTA TTTTGATAAA TATGTCGACC
3451  CCAGCTAGCG ATTGCTAAGT ATTGGATGTT TGAAACGTCC ATGACCCAAT
3501  CGATTAATTG GTTGAACTCG TCAATGGTAT ATCCTAAATC ACTGCCTCTG
3551  GTTTCTTCTG GCGTTGCTCG AAATCCTAAA ATAAAATTGT CAGGTGCTTC
3601  TTTATCAATC ACTTCTTGTA CCGCACGCAT AACTTCTAAA CATAATCTTG
3651  CACGATTTTT TAATGAGTCG GCACCGTAAT GGTCTGTACG TCTATTTGAA
3701  AAAGTTGAGA AAAATGTTTG AATCAGCAAA CGTTGTGCAA TCGAAATTTC
3751  CACACCATCA AAACCTGCTT TAATCGCGCG TGCATCGAGC TCGTGCC
```

SEQUENCE 3 [SEQ ID NO:17]
gactaataat actgaacg

SEQUENCE 4 [SEQ ID NO:18]
tctgtcggtt tctctggg

Gene #7
Fatty Acid Oxidation Complex Subunit

SEQUENCE 1 [SEQ ID NO:19]

```
1  CAGGCGTTTC CTCNGGTACN TGTTGCNNGC CTTTAATTAC CGACNCTGCA
51  ATANCCAAAC CGACCAGGTC GGATAGGGNA TATGTACCTG TTTTAGGACG
101  ACCAATCGCT TGCCCAGTTA AAGCATCCAC ATCTACNATG CTTANCTTGT
151  GTTGCTCGGC GCGATACAGA ATATCATTCA TTGTGTGCGT GCCGACTCTA
201  TTTGCGACAA AGCCAGGCAC ATCATTGACG ACAATGACAC CTTTACCTAA
251  TACATTGTGC GCGAAATTTT TTACATCTAA TATGATAGAT TCCTTCGTGT
301  GTGACGTAGG TATTAACTCC ACTAATTNCA TAATACGTGG TGGGTTAAAG
351  AAATGTAGAC CAAAGAATCG CTCTTGATCC TTCTCGTTAA ATGCTTGAGC
```

401 AATCGCATTA ATTGGGATTA CCTGATGTAT TTGTAGCAAA TAAAGCATCT

451 TCTNTAGCAT GTTGTAGAAC TTGTTGCCAA ACAGCATGCT TAATTTCAAT

501 ATCTTCTTTG ACTGCTTCGA TATATAAATC AGNATCATCA TTTACCAAGT

551 CATCATCAAA ATTACCATAT GTTAAATGAC TCACTAGATT TAAGTCGAAT

601 AGTAGCGGCC GTTTCTTATC TGTAATTTTA TCGTAAGATT TTTTCGCAAT

651 GAGATTTGGA TCGTTTGTGT CCACTACAAT ATCTAATAGT TTTACTTTAA

701 GTCCAGCATN CACAAAGAGT GCTGCCAGTT GAGCGCCCAT CGTGCCTGCG

751 CCAAGAACGG TTACTTTATT AATTGTCATA GTGATTCCTC CAATTTAGGT

801 GAGGATAAGA TAACCATTAA GATAATTGGA ATAACGNTGC TATTTTATNA

851 AATTAATTAA GTATCTTTGA CAAGACATCT CAGNCTCTTT ATTTTAAGGA

901 AAAAGCTTTA TGCTTAAAAT AAGTCTTTTT TAGTGAAATT AATGCATCTC

951 ATATAATTAT TTGCTATTTA TACGAAAGCA GAATCTCCAG TCAAAGCGCG

1001 TCCAATTACT AAGGCATTAA TTTCATGTGT ACCTTCGTAC GTGTAAATCG

1051 CTTCTGCATC AGAGAAGAAA CGTGCAATAT CATAATCGTC AGCTAGTATG

1101 CCATTACCAC CTGTAATACC GCGGCCCATA GCTACTGTCT CACGCAAACG

1151 TAAGGCATTC ATCATCTTCG CCGGTGAAGT TGCAACCTCG TCATATTCAC

1201 CATGTGCTTG CATATTAGCT AATTGAGCAC ATGTTGCCAT TGCTTGAGCT

1251 AAATTACCTT GCATCATTGC TAGCTTNTCT TGTATTAACT GATATTTACT

1301 AATTGGGTNT GCCGAATTGC TTACGCTCAA GTGACATAAT CTAATGTGGC

1351 ACGTAAAGCG CCAGCCATAC CACCTGTAGC CATATAAGCA ACGCCTGCTC

1401 TCCGGTGGAA TAAAGAATTT TG

SEQUENCE 2 [SEQ ID NO:83]

1 MLXKMLYLLQ IHQVIPINAI AQAFNEKDQE RFFGLHFFNP PRIMXLVELI

51 PTSHTKESII LDVKNFAHNV LGKGVIVVND VPGFVANRVG THTMNDILYR

101 AEQHKXSXVD VDALTGQAIG RPKTGTYXLS DLVGLXIAXS VIKGXQXVPE

151 ETP

SEQUENCE 3 [SEQ ID NO:20]

atgtacctgt tttaggac

SEQUENCE 4 [SEQ ID NO:21]
gagtcattta acatatgg

Gene #8
ATP DEPENDENT RNA HELICASE DEAD

SEQUENCE 1 [SEQ ID NO:22]

```
       1  ATACTTTGAT TTTAGATGAA GCTGATGAAA TGATGAATAT GGGATTCATC
      51  GATGATATGA GATTTATTAT GGATAAAATT CCAGCAGTAC AACGTCAAAC
     101  AATGTTGTTC TCAGCTACAA TGCCTAAAGC AATCCAAGCT TTAGTACAAC
     151  AATTTATGAA ATCACCAAAA ATCATTAAGA CAATGAATAA TGAAATGTCT
     201  GATCCACAAA TCGAAGAATT CTATACAATT GTTAAAGAAT TAGAGAAATT
     251  TGATACATTT ACAAATTTCC TAGATGTTCA TCAACCTGAA TTAGCAATCG
     301  TATTCGGACG TACAAAACGT CGTGTTGATG AATTAACAAG TGCTTTGATT
     351  TCTAAAGGAT ATAAAGCTGA AGGCTTACAT GGTGATATTA CACAAGCGAA
     401  ACGTTTAGAA GTATTAAAGA AATTTAAAAA TGACCAAATT AATATTTTAG
     451  TCGCTACTGA TGTAGCAGCA AGAGGACTAG ATATTTCTGG TGTGAGTCAT
     501  GTTTATAACT TTGATATACC TCAAGATACT GAAAGCTATA CACACCGTAT
     551  TGGTCGTACG GGTCGGTGCT GGTAAAGAAG GTATCGCTTG TAACGTTTGG
     601  TTAATCCAAT CGAAATGGAT TATATCAAGA CAAATTGAAG ATGCAAACGG
     651  GTAGAAAAAT GAGTGACTCC GCCACCTCAT CGGTAAGAAG TACTTCCAAG
     701  CACGTGAGGA TGACATCAAA GGAAAAGGTG GAAACTGGAT GTCTTTAAGA
     751  GTCAAGAATC ACGCTGGAAA CGCATTCTTC AGAGGTGGGT AAATTGAATT
     801  TTACGATGTG G
```

SEQUENCE 3 [SEQ ID NO:23]
gatgaagctg atgaaatg

SEQUENCE 4 [SEQ ID NO:24]
tatctagtcc tcttgctg

Gene #9
PHOSPHORIBOSYLAMINE GLYCINE LIGASE

SEQUENCE 1 [SEQ ID NO: 25]

```
       1  TAATTCGCAA TAGGAGTGAT GAATATCATA AATTTTACCC TCCAAATGAA
```

```
      51  GCTAATGAAG TCCTGGACCC GAGTAAGACG CATGTAGCCA AGCTAAAATA
     101  ATCCACTCTA CCTTATCTTT AGTTAATAAT GTTACTAAAT GTTGTTCATA
     151  CGCTGCTTTT GAATCAAATT GTTTTGGTTC ATTAATATAA ACAGGAATAT
     201  CGTGCTTGTT TGCTCTATCT ATACAAAACG CATTTTGATG ATCCGTATAT
     251  AGCNCCGTAA CTTCAATATT TTCAAGTTTT CCTGATTCAA CATGCTCAAC
     301  TATATTTTCA AAGTTACTTC CTGAACCTGA TGCAAAAATC GCAATTTTAA
     351  CCATTGTTAT ACCCCCAACA ATTCAATTGC AGTTGACTCA TTTTTCACAA
     401  TATGACCAAT TTGATAAGCT TCCACATTTT GTTCTGCTAA AATCTTCAAA
     451  GCGCGTCGAT GCATCTTTTT CATCAACGAT AACCGTATAG CCAATACCCA
     501  TGTTAAAAAT GTTATACATT TCATTTGTGT CTATATTGCC TTGTTGTTGT
     551  AACCAATCAA ATATTTTTGG CGTTGGAAAT GATGTAGTAT CAATTCTAGC
     601  AGCATATCCG GCTGGCAATG CACGTGGAAT ATTTTCATAA AAACCTCCAC
     651  CAGTAATATG ATTCATTGCC TTAATAGAAA CTTCTTTTTT TAAAGCAAGT
     701  ACAGGTNTGA CATATAATTT AGTTGGCTCT AAAAAGACAT CTATAAATGG
     751  ACGATTATCG NAGGGTGATG CCAAATCAAT GNCTGATTCA NTAATTAATN
     801  TGCGCACTAA ACTGTNTCCA TTNGANTGAA TGNCACTTGG ACGCAAGTCC
     851  TATAACAACT TGGCCCTCTT NCAATTCTTG AACCATCTTA CAATAGNCAA
     901  CCTTTTTCAA CTGCTCCAAC AGCAAATCCG GCTACATCAT ATTCACCTTC
     951  GTGATACATT
```

SEQUENCE 3 [SEQ ID NO: 26]
ataagcttcc acattttg

SEQUENCE 4 [SEQ ID NO: 27]
gataatcgtc catttata

Gene #10
Methanobacteria formate dehydrogenase

SEQUENCE 1 [SEQ ID NO: 28]

```
       1  GGCACGAGCG CTAAATAATT AATATTTAGT TTTTAAGTTA TTAATAACGT
      51  AGGGATATTA ATTTTAAAAG AAGCAGACAA AATGGTGTTT GCTTCTTTTT
     101  TATGTCGTAT AAGTAATAAA TAAAACAGTT TGATTTTAAA ATGAAAGCGT
```

151 AAAAATGGTA AAATATCCCA AATTGATTG TGATATAATT ATAAGGAAAA

201 TGAGCAATTT ATGAAAAAAG TTTACGNACA AATCGGAGAA TTAAAACTAA

251 ATAATTATCA AAACAACGTC AATATTTAGT TGAATACTCA GACTTTAGCC

301 CATGGCCAAG TGGGGAAGAC AGCATATATT AGTAAAGGTG AATGATTTGT

351 TATTACTCAC TCGAAAATAG AAAGACAAGA TTTTAACGAT TAAAATAAAC

401 TATTTTACAA ATAAAGTAAA ATTAATTTAT TANGCTAATA ATGCAAAAAA

451 TTAAAAAGTA ATGGACAAAG AGATAATGAT ATGGCTCAAG AGGTAATAAA

501 ATAGAGGTGG ACGCACACTA AATGGGGAAG TTAATACAAG G

SEQUENCE 3 [SEQ ID NO: 29]
gcacgagcgc taaatttg

SEQUENCE 4 [SEQ ID NO: 30]
CTTCCCCATT TAGTGTGC

Gene #11
E.coli Nitrate Reductase

SEQUENCE 1 [SEQ ID NO: 31]

1 CCACCCANCT GATTATAATG TTTTAGCANG AGCTAGACTT GGTTGGTTAC

51 CATCATATCC ACAATTTAAT AAAAATAGTT TGTTGTTTGC AGAAGAAGCT

101 AAAGATGAAG GCATTGAGTC GAATGAGGCA ATTTTAAAAC GAGCGATAAA

151 TGGAAGTTAA GTCAAAACAA ACGCAATTTG CGATAGAAGA TCCGGATTTG

201 AAAAAGAATC ATCCGGAAAT CACTGTTTAT ATGGCGCTCA AATCTAATCT

251 CAAGTTCTGC AAAAGGTCAA GAATACTTTA TGAAGCATTT ACTTGGCACA

301 AAATCAGGGT TATTAGCTAC ACCAAATGAA GATGAAAAGC CAGAAGAAAT

351 TACGTGGCGT GAGGAAACAA CAGGGAAATT AGATTTAGTC GTTTCTTTAG

401 ATTTCAGAAT GACAGCAACA CCTTTATATT CTGACATTGT TTTGCCAGCA

451 GCGACTTGGT ATGAGAAGCA TGATTTGTCA TCTACAGATA TGCATCCATA

501 TGTACATCCT TTTAATCCAG CTATTGATCC ATTATGGGAA TCGCGTTCAG

551 ACTGGGATAT TTATAAAACG TTGGCAAAAG CATTTTCAGA AATGGCAAAA

601 GACTATTTAC CTGGAACGTT TAAAGATGTT GTGACAACTC CACTTAGTCA

651 TGATACAAAG CAAGAAATTT CAACACCATA CGGCGTAGTG AAAGATTGGT

701 CGAAGGGTGA AATTGAAGCG GTACCTGGAC GTACAATGCC TAACTTTGCA

751 ATTGTAGAAC GCGACTACAC TAAAATTTAC GACAAATATG TCACGCTTGG

801 TCCTGTACTT GAAAAAGGGA AAGTTGGAGC ACATGGTGTA AGTTTCGGTG

851 TCAGTGAACA ATATGAAGAA TTAAAAAGTA TGTTAGGTAC GTGGAGTGAT

901 ACAAATGATG ATTCTGTGAG AGCGAATCGT CCGCGTATTG ATACAGCACG

951 TAATGTAGCA GATGCAATAC TAAGTATTTC ATCTGCTACG AATGGTAAAT

1001 TATCACAAAA ATCATATGAA GATCTTGAAG AACAAACTGG AATGCCGTTA

1051 AAAGATATTT CTAGCGAACG TGCTGCTGAG AAAATTCGTT TTTAAATATA

1101 ACTTCACAAC CACGAGAAGT AATACCGACA GCAGTATTCC CAGGTTCAAA

1151 TAAACAAGGT CGACGATATT CACCATTTAC AACGAATATA GAACGTCTAG

1201 TACCTTTTAG AACATTAACA GGACGTCAAA GTTATTATGT GGATCACGAA

1251 GTTTTCCAAC AATTTGGGGA GAGCTTACCA GTATATAAAC CGACATTGCC

1301 GCCAATGGTA TTTGGGAATA GAGATAAGAA AATTAANGGT GGTACAGATG

1351 CTTTGGTACT GCGTTATTTA ACGCCTCATG GANAATGGAA TATACACTCA

1401 ATGTATCAAG ATAATAAGCA TATGTTGACA CTATTTAGAG GTGTCCACCG

1451 GTTTGGATAT CANATGAAGA TGCTGNAAAA CACGATATCC AAGATAATGA

1501 TTGGCTAGAA GTGTATANCC GTAATGGTGT TGTAACGGCA AGAGCAGTTA

1551 TTTCGCATCG TATGCCTAAA GGTACAATGT TTATGTATCA TGCACAAGAT

1601 AAACATATTC AAACGCCTGG GTCAGAAATT ACAGATACAC GTGGTGGTTC

1651 ACACAACGCG CCGACTAGAA TCCATTTGAA ACCAACACAA CTAGTCGGAG

1701 GATACGCACA AATTAGTTAT CACTTTAATT ATTATGGACC AATTGGGAAC

1751 CAAAGGGATT TATATGTAGC AGTTAGAAAG ATGAAGGAGG TTAATTGGCT

1801 TGAAGATTAA AGCGCAAGTT GCGATGGTAT TAAATTTAGA TAAATGCATA

1851 GGATGCCATA CGTGTAGTGT GACATGTAAA AACACTTGGA CAAATCGTCC

1901 AGGTGCTGAG TAACATGTGG TTCAATAACG TAGAAACGAA GCCAGGTGTA

1951 GGGTATCCGA AACGTTGGGA AGACCAAGAA CACTACAAAG GTGGTTGGGT

2001 ACTAAANTCG TAAAGGGAAA CTTGAATTAA AATCTGGAAG TAGAATTTCA

2051 CAAATTGCTT TAGGTAAAAT TTTTTATAAC CCAGATATNC CATTAATAAA

2101 AGATTATTAT GANCCATGGA NCTATAATTA TGAACATTTA ACAACTGCGA

2151 AATCAGGGAA GCATTCGCCA GTTGCTAGAG CGTATTCAGA AATTACAGGG

2201 GATAACATTG AAATTGAATG GGGACCTAAC TGGGAAGATG ACTTAGCAGG

2251 TGGTCATGTT ACAGGCCCAA AAGATCCTAA CATACACAAA ATAGAAGAAG

2301 AGATTAAATT CCAATTTGAC GAAACTTTTA TGAG

SEQUENCE 2 [SEQ ID NO:84]

1 MKHLLGTKSG LLATPNEDEK PEEITWREET TGKLDLVVSL DFRMTATPLY

51 SDIVLPAATW YEKHDLSSTD MHPYVHPFNP AIDPLWESRS DWDIYKTLAK

101 AFSEMAKDYL PGTFKDVVTT PLSHDTKQEI STPYGVVKDW SKGEIEAVPG

151 RTMPNFAIVE RDYTKIYDKY VTLGPVLEKG KVGAHGVSFG VSEQYEELKS

201 MLGTWSDTND DSVRANRPRI DTARNVADAI LSISSATNGK LSQKSYEDLE

251 EQTGMPLKDI SSERAAEKIR F*

SEQUENCE 3 [SEQ ID NO: 32]
attgatccat tatgggaa

SEQUENCE 4 [SEQ ID NO: 33]
catattgttc actgacac

Gene #12
E.coli ftsE (abc transporter)

SEQUENCE 1 [SEQ ID NO: 34]

1 AGTTATTGTA TTTAAAAATG TTTCATTTCA ATATCAAAGT GATGCATCCT

51 TCACATTGAA AGATGTTTCT TTTAATATAC CTAAAGGTCA GTGGACATCT

101 ATTGTTGGTC ATAACGGTTC TGGAAAATCT ACAATTGNCA AGTTAATGAT

151 TGGCATAGAG AAAGTTAAAT CTGGAGAAAT TTTTTATAAT AATCAAGCTA

201 TAACTGATGA TAATTNTGAA AAGTTAAGAA AAGACATAGG AATTGTATNT

251 CAGAATCCGG ATAATCAATN TGTTGGNTCA ATTGTAAAAT ACGATGTGGC

301 ATTTGGACTC GAAAATCATG CGGNTCCACA TGACGAAATG CATAGAAGAG

351 TCAGCGAAGC ACTTAAACAA GTTGATATGT TAGAACGTGC AGATTATGAC

401 CCTAATGCAT TATCGGGGGG ACAGAAGCAG CGTGTGGCTA TAGCAAGTGT

451 ATTAGCACTT AACCCTCTGT CATTATATAG ATGAGGCGAC TCTATGTTAG

501 GATCCCTGAT GCACGTCAAA TTTATGGGAT TTAGNGAGAA AGTAANTCAG

551 ACATTATATA CAATCATTCT ATACGCATGA TTTATCTGAG GCGATGAGNA

601 GATCAAGTAT CCGTATGATA AGGACTTNCT TTTAAGGC

SEQUENCE 3 [SEQ ID NO: 35]
gtttcatttc aatatcaa

SEQUENCE 4 [SEQ ID NO: 36]
atctatataa tgacagag

Gene #13
B.subtilis secA

SEQUENCE 1 [SEQ ID NO: 37]

1 GTTAATCAAG TATCGAAGCG GAACAATCAT ACTTTAATGT TGAAGATTTA

51 TATNGCGAAC AAGCGATGGT CCTAGTGCGT AATATTAATT TAGCACTGCG

101 CGCACAATAT TTGTTNGNAT CTNATGTCGA TTACTTTGTA TATNNTGGTG

151 ATATTGTTTT AACTGACCNC ATTACAGGTC GTNTGTTACC GGNAACTAAG

201 TTGCAAGCTG GACTTCACCA NGCTATTGAA GCGAAAGAAG GTATGGAGGT

251 TTCAACAGAT AAAAGTGTTA TGCCAACCAA TTACCCTTCC AGAATTTATT

301 TAAACTTTTT GAATCAATTT TCAGGTATGA CAAGCTACAG GAAAATTAGG

351 CGAATCAGAG TTCTTTGATT TGTATTCANA AATAGTCGTA CAAGCACCCA

401 ACTGATAAAG CGATTCAACG TATCGATGAA CCAGATAAAG TGTTTCGTTC

451 AGTTGATGAG AAAAACATCG CGATGATTCA TTGATATAGT TGAACTTCAT

501 GANNCGGGGC CGACCGGTTT TACCTCATAA CCGAGNACTG CTGAAGCGGC

551 TTGAATACTT TTCNGAAGTA TTATTCCAAA TGGATATTCC TAATAATTTA

601 CTCATTGCGC AAAATGTTCC AAAAGAAGCG CAGATGATAG CTGAAGCAGG

651 CCAAATTGGT TCCATGACTG TTGCGACTAG TATGGCAGGT CGAGGCACAG

701 ATATTAAACT TGGTGAAGGT GTCGAAGCAT TAGCTGGATT AGCTGTTATT

751 ATTCATGAAC ATATGGAAAA TAGCCGTGTA GACAGGCAAT TACGTGGTCG

801 TTCTGGTAGA CAAGGGGATC CGGGATCATC TTGTATATAT ATTTCACTAG

851 ATGATTATTT AGNTAAGCGA TGGAGCGATA GTAATTTAGC GGAAAATAAT

901 CAATTATATT CANTAGATGC ACAACGATTA TCGCAAAGTA ATTTGTTTAA

951 TCGNAAAGTT AAGCAAATTG TAGTTAAAGC GCAGCGTATC TCGGAAAGAA

1001 CAAGGGGTTA AAGCTCGGTG AAATGGCTTA ATTGAATTTG NNAAAAAGCA

1051 TNAGTATTCA GCGAAGATCT TNGTATTTAC GANGGAACGC AAATCCGAGT

1101 TTTTAGAAAT TAGATTGATG CTGAGAATCC NAGATTTTTA ANGCGGTTAG

1151 CTTAAAGATT GTATTTGAAA TNGTTTGGGG NAATGANGGA AANGGTGCTA

1201 ACAAAATCGC GNGTTGGGCG AGTATATTTT ATCAAAAATT TAAGTTNCCA

1251 ATTTAATAAA GATGTGGCTT GTGTTAATTT TAAAGATAAG CAAGCAGNAG

1301 TGACATTTTT ATTAGAGCAA TTTGAAAAGC AATTAGCTTT GGANTCCGTA

1351 AAAACATGCA ANGNGCATAT TATTATAATA TTNCCGGCCA AAANGTCTTT

1401 NGGGAAAGCA ATTGATNCAA GTTGGGGTTA GGAACAAGTC GGCTTTTNAC

1451 AACAANTTAA NAGCAAGCGN TAATCAAACG ACAAAANTGG CAACCT

SEQUENCE 2 [SEQ ID NO:85]

1 MDIPNNLLIA QNVPKEAQMI AEAGQIGSMT VATSMAGRGT DIKLGEGVEA

51 LAGLAVIIHE HMENSRVDRQ LRGRSGRQGD PGSSCIYISL DDYLXKRWSD

101 SNLAENNQLY SXDAQRLSQS NLFNRKVKQI VVKAQRISER TRG*

SEQUENCE 3 [SEQ ID NO: 38]
ccgctaaatt actatcgc

SEQUENCE 4 [SEQ ID NO: 39]
ctgaagcggc ttgaatac

Gene #14
E.coli choline dehydrogenase

SEQUENCE 1 [SEQ ID NO: 40]

1 ATATAAATTA TTTAAGCGTA TGGTTTTACT TCGATTGCAC CCTTCATTTT

51 CATCATTGAA CACCATGCTT AATATAATCC ATATATTTGT GGCTCTAAAG

101 NCTTTCCTCC CACCGTATAA TGTCTGCTGC TTTTTCAGCT AACATTAAAA

151 CAGGTGCGTG TATATTGCCA TTTGTCGTAC GTGGCATAGC GGATGCATCA

201 ACTACACGTA AATTTTCCAT ACCGTGGACT TTCATTGTTA ACGGGTCAAC

251 TACTGCCATT GGATNCTGAA GCAGGACCCA TTTTAGCACN ACAAGATGGG

301 TGTAATNCTG TTTCACCATC TCNACGGAAN NCAATCAAGN ATTTCTTCGT

```
 351  CTGTTTGCAC TTCTGGGTCC TGGGTGAAAT TTCTCCACCA TTGAATGGAT
 401  CCATTGCTTT TTGAGATAAG ATATTTCTTG CTACACGAAT TGCTTCTACC
 451  CATTCTNTTT TATCTTCTTC TGTTGATAAA TAATTAAAGC GGATACTTGG
 501  TTTTTCGAAT GGATCTTTAG ATTTGATTGG CACGAGCTAC CACGAGAGTT
 551  TGAATACATT GGTCCTACGT GAACTTGATA ACCATGTGCG ACCGCTGCCT
 601  TTTGACCATC ATATCTTACA NCTATTGGTA AGAAATGGAA CATTAAGTTA
 651  GGATAATCAA CTTCGTTATT TGAACGTACA AATCCGCCAC CTTCAAAATG
 701  GTTAGATGCT GCTGCACCTG TACGTGTGAA AATCCAGTGG TAAACCAATT
 751  AAATGGCATG CGCCTTGATA TCTAAGCTTG GCTGTAATGA TACAGGTTTC
 801  CTTACATTTA TGTTGAATGT ATACCTCTAA GTGATCTTCC AAAGTTTTCA
 851  CCCACACCTG GTAAATGAAC ACGTGGCTCA ATGCCTTTTG ATTTTAGGAA
 901  CTCTGAATCA CCGATACCAG ATAATTGTAG TAATTGTGGC GTTATTGAAT
 951  GCCCC
```

SEQUENCE 3 [SEQ ID NO: 41]
gaagcaggac ccatttta

SEQUENCE 4 [SEQ ID NO: 42]
gattttcaca cgtacagg

Gene #15
S.aureus DNA Gyrase

SEQUENCE 1 [SEQ ID NO: 43]

```
   1  GAATTCCTAC ATAATACTTT TGTTTACCTT GTGTCAGTTT ATACAACGGT
  51  GGCTGTGCAA TATACACATA GCCTGCTTCA ATTAACGGTC TCATAAATCG
 101  ATAGAAGAAT GTTAATAACA ATGTTCTAAT ATGCGCTCCA TCCACATCGG
 151  CATCAGTCAT AATGACGATT TTGTGATATC TTGCTTTCGC TAGATCAAAG
 201  TCGCCACCGA TTCCTGTACC AAATGCTGTG ATCATTTGAC GAATTTCATT
 251  GTTATTCAAA ATTCTATCTA ATCGTGCTTT NTCAACATTT AATATCTTAC
 301  CTCGTAATGG TAAAATCGCC TGCGTTCTAG AGTCACGACA GATTTTGGTG
 351  GACCCCCNGC AGAGTCCCCT TCGACTAAGA AAATCTCACA TTCTTCAGGA
 401  CTTTTACTAG AGCAATCGGC TAATTTACTG GAAGACTGCT ACATCTACGC
```

451 TGATTTACGA GGTGTTACTT CAGGGCTTTN TCGAGACACG TGCANGT

SEQUENCE 3 [SEQ ID NO: 44]
cataatactt ttgtttacc

SEQUENCE 4 [SEQ ID NO: 45]
agtaacacct cgtaaatc

Gene #16
E.coli pts system ptkC

SEQUENCE 1 [SEQ ID NO: 46]

1 CTANCNAANG GAANTTCAGC ATCCTTAAAA ATACCTATTT GACTGTAGAA

51 ACCTTTTGNT GCGTACAATA TCTAAACCTT GTCGTGCTGC TGGAACTGCA

101 CCTGAACATT CAACAACAAC ATCTGCACCG TAACCGTCTG TAATTCCATT

151 GATATACGTT TTTAAGTCTG TGTGTTGTAA ATTGACTACA TAATCCATGT

201 GCAATGCTTC TGCTTTATCT AATCTGACTT NGTGGCANTG TCCAATCCAG

251 TTACCACAAC AGGTGCGCCT TTACTTTTCA ACACTTGTGC TACAAGTAAT

301 CCGATTGGCC CAGGTCCCAT TACAACTGCT ACATCGCCAG AGTTCACTTG

351 AATCTTAGAA ACGCCATGAT GTGCACATGC TAATGGTTCT TGTCATAGCT

401 GCAGACTGAT ACGATACTTC CGCTTCTGGA ATATGATNCA AACTTTCTTC

451 ACGTGCAATG ACATAATTAG TAAATGCGCC ATCAACTTGT GTTCCAATAC

501 CTTTTCGATG GTTGCATAAA TGATAGTTTT TTGATTTACA GGAATCACAC

551 TCATTACANA CCATAGAATG TAGTTTCAGA AGTGACNCGG TCACCAACTT

601 TAAAATCNTT AACGTCTGCT CCCAACTTCA ACGATNTCAC CAGAAAATTC

651 ATGACCTAAT GTCACTGGAA AATTAACTTN ATAATGCCCT TCATAAGTAT

701 GAAGGTCTGT GCCACAAATT CCTGCATAAT GTACTTTAAT CTTTACTTTA

751 TCATCTAGCG GTGTTGCAAC TTCTTTATCA AGAAGTTCTA AGTTGCCATG

801 TCCTTCTCTT GTTTTTACTA AAGCTTCCAC CACAAACACN TCGANTTTTT

851 ANTTGNAATA GACTNNATAG NTTNAAGATA AGATAGTTAN CGATATTNCC

901 ACCTTGATCA ATACTTGANA TTTCAGATGA ACCTTTTGNC ATTTGTACAT

951 TCGTACCTTT CGCCATATCT GTGAAAATGG GTGCTACGTC TGTTGCAATA

1001 TATAATGAAA TTGCAATCAT AATCGTACCC ACAATGACAG AATGAATAAT

1051 GTTTCCTCTT GCTGCACCAA CAATAAACGC GACAACAAAT GGTATAGTTG

1101 CTAAGTCACC AAAAGGTAGT ACTTGGTTTC CTGGTAAAAT AACGGCTAAT

1151 AAAACAGTGA TAGGTACTAA AATTAATGCT GTCGAAATAA CCGCTGGATG

1201 ACCTAATGCT ACAGCCGCAT CCAATCCAAT ATAAATTTCA CGTTCGCCAA

1251 AACGTTTATT TAGCCATGTT CTTGCAGACT CTGAAACTGG CATTAAACCT

1301 TCCATTAAGA TTTTTACCAT TCTAGGCATT AAGACCATTA CTGCAGCCAT

1351 TGACATTCCT AAATTAATGA TGTCTCCAGG TTTGTAACCT GCTAACACAC

1401 CAATACCTAA ACCTAAAATT AAGCCGACAA ATATAGACTC TCC

SEQUENCE 2 [SEQ ID NO:86]

1 GESIFVGLIL GLGIGVLAGY KPGDIINLGM SMAAVMVLMP RMVKILMEGL

51 MPVSESARTW LNKRFGEREI YIGLDAAVAL GHPAVISTAL ILVPITVLLA

101 VILPGNQVLP FGDLATIPFV VAFIVGAARG NIIHSVIVGT IMIAISLYIA

151 TDVAPIFTDM AKGTNVQMXK GSSEXSSIDQ GGNIXNYLIX XLXSLXQXKX

201 RXVCGGSFSK NKRRTWQLRT S*

SEQUENCE 3 [SEQ ID NO: 47]
gttctaagtt gccatgtc

SEQUENCE 4 [SEQ ID NO: 48]
cctagaatgg taaaaatc

Gene #17
S.typhimurium adenine glycosylase

SEQUENCE 1 [SEQ ID NO: 49]

1 CCATTTAAAA GTATTGTAAA ATCATCCACN TTNTATAAAC CAACCACNTT

51 AACNTTTTTG ACATTTGTTA TCCGATGAGA TTAAAAGATA TCAATNAATA

101 CAATTTTTAN AATTAATGTC ACTATGTTTT CCGATAATAT NACCCAATCA

151 TCGNAATGTT ACCCATTTAT AAAATGANAA ATCNTTGACA TAGGTANAGG

201 GAATGTATAT TGGTCNCGGA TCACTTAAAT TAAACCCANA TCATGTCATC

251 TGGTAATGTN TCAATGTTAA TTGCTCCTGA AGCGGCGTAN ACTTTAATCT

301 TCCATGTTAA ATGAGTAAAT TGATGCGTCA ACTCNAAAAT AGGTGTTTCT

351 NCTGGNTGAA TGTCATGACC GATTTTTTCA NTCATTTTAC GTCTANCATG

401 CTCACTATCN AACATAGGAN ATTGCCACAT ACCATACNAT AATTNTTCCC

451 TACGCTTTTG CAACAGATAT TGACCTTGAT TATTTCTAAT TAANAAGACG

501 GATTGCTCAA TTACNTTTTT ACTTACATTT TTAGATTTAA CAGGTAACTT

551 TTCAAATGGA CCTTTATCAA ATGCCTCACA GTTTTCTTGN ACTGGACNAA

601 ATAAGCATAA TGGATTTTTT GGTGNACAAA TTAATGCCCC TAATTCCATC

651 ATAGCTTGAT TAAACGTTCC AGCTTCTGTA GTAACATACG GTAACAATTC

701 TTGTTCGTAC GATTTCCTCG TCGATTGTAA TTTAATATCT CGATAGTCAT

751 CATTCAATCT AGACCATACG CGAAAAACAT TTCCGTCTAC AGTTGCTAGT

801 GGTACATTAT ATGCAATGCT CATTACTGCA GCTTGTGTGT ATGGGCCAAC

851 ACCTTTTAAC GCTTTAAATT GATCAGGATC TTTGGGAACT AAGCCTTCAT

901 ATTTATCANA AACTTCTTTA ATCGCCGTAT GAAAATTTCG AGCTCTACTA

951 TAATATCCTA AGCCTTCCCA ATACTTTAAC ACTTCATCTT CCGAAGCTTG

1001 ACTCAAAACT TCCACAGTTG GAAATCGGNC ACCAAAACGA TGATAATAGT

1051 CAATAACTGT TTTAACTTGT GTCTGTTGTA ACATGACCTC ACTTAACCAA

1101 ATATAGTACG GATTGGTCGT TTGTCGCCAT GGCATTTCTC TTTGATTTTC

1151 ATCAAACCAG TGTATCAAAT TTTCTTTAAA ACTAGACTGC TGATACATTT

1201 ATAAAACCCT TTCCTCACCA AAATTAATTG TCTTTACTCA TAATGTTTTT

1251 ATTGTACATT AAAATCATGG TTAGTATGTA AGTTAATTTA GTTATNTGCG

1301 AAATTGGATT ATAATAGTAT ATATAATATT ATGAAATGAG TGAACTGATA

1351 TGGACACTGC AACACATATC GCAATTGGGG TGGGCCTTAC AGCACTTGCA

1401 ACTCAAGATC CAGCAATGGC TTCTACGTTT GGTGCAACAG CTACAACCCT

1451 TATCGTTGGT TCATTAATTC CTGATGGGGA TANTGTNCTT AAATTANAGG

1501 ACANTGCAAC ATATATTTCG NATCATAGAG GNATNACGTC ATNCCATCCC

1551 CTCCCACAAN NNTATGNCCA GTCNCNTTTA CANTTTNTAT NTNTTCACGT

1601 CACTNTNGCT GGTANGCATC CCNCCTCACG TATGGCTTGT GG

SEQUENCE 2 [SEQ ID NO:87]

1 MYQQSSFKEN LIHWFDENQR EMPWRQTTNP YYIWLSEVML QQTQVKTVID

51 YYHRFGXRFP TVEVLSQASE DEVLKYWEGL GYYSRARNFH TAIKEVXDKY

101 EGLVPKDPDQ FKALKGVGPY TQAAVMSIAY NVPLATVDGN VFRVWSRLND

151 DYRDIKLQST RKSYEQELLP YVTTEAGTFN QAMMELGALI CXPKNPLCLF

201 XPVQENCEAF DKGPFEKLPV KSKNVSKXVI EQSVXLIRNN QGQYLLQKRR

251 EXLXYGMWQX PMXDSEHXRR KMXEKIGHDI XPXETPIXEL THQFTHLTWK

301 IKVYAASGAI NIXTLPDDMX WV*

SEQUENCE 3 [SEQ ID NO: 50]
tcctgaagcg gcgtatac

SEQUENCE 4 [SEQ ID NO: 51]
tatgaaggct tagttccc

Gene #18
S.aureus femA

SEQUENCE 1 [SEQ ID NO: 52]

1 GGGAAAAAAA GAAAACCTTC CAAAATACGG GAAATTGAAA TTAATTANCC

51 GGAGAGACCA NATAGGAAGT AATTGATAAT GGAAGTTTCC CCANAATTTA

101 ACAAGCTAAA AGAGTTTGGG TGCCTTTTAC AAGATAAGCA TGCCAATACA

151 GTCATTTCAC GCACACTGTT GNCCACTATG AGTTAAAGCT TGCTGAAGGT

201 TATGAAACAC ATTTAGTGGG AATAAAAAAC AATAATAACG AGGTCATTGC

251 AGCTTGCTTA CTTACTGCTG TACCTGTTAT GAAAGTGTTC AAGTATTTTT

301 ATTCAAATCG CGGTCCAGTG ATCGATTATG AAAATCAAGA ACTCGTACAC

351 TTTTTCTTTA ATGAATTATC ANAATATGTT AAAAAACATC GTTGTCTATA

401 CCTACATATC GATCCATATT TACCATATCA ATACTTGAAT CATGATGGCG

451 AGATTACAGG TAAGGCTGGT AATGATTGGT TCTTTGATAA AATGAGTAAC

501 TTAGGATTTG AACG

SEQUENCE 3 [SEQ ID NO: 53]
gaggtcattg cagcttgc

SEQUENCE 4 [SEQ ID NO: 54]
CAAATCCTAA GTTACTCATT

Gene #19
Parsley S-adenosyl methionine synthetase

SEQUENCE 1 [SEQ ID NO: 55]

1 CGCACATAAC GTGCAGCATA TGCAGCTGAG CGGTCTACTT TTTGTAGGAT

51 CCTTACCACT GAAGCATCCG CCACCATGAC GTGCATAGCC ACCATACGTA

101 TCAACAATGA TTTTACGTCC TGTTAATCCT GCATCACCTT GAGGTCCACC

151 GATTACAAAG CGTCCTGTAG GATTGATGTA GAATTTAGTT TGTTCATTAA

201 TCAAGTTTTC TGGAACAGTT GGATAAATGA CATGCGCTTT GATGTCTTCT

251 TGAATTTGTT CAAGTGTCAC ATCATCAGCA TGTTGTGTTG ATACGACAAT

301 CGTATCAATA CGTACTGGGT TATCATTTTC ATCATATTCA ACAGTGACCT

351 GAACTTTACC GTCTGGTCGT AAATAATTCA ACGTCTCGNG CCATCTTTTA

401 CGCACATCAG ATTAAACGTT TGGGGCAATT GGGTGTGATA AATTAAATTG

451 CTAGAGGGAT GTACGTTTCT TGTTTCAAT

SEQUENCE 3 [SEQ ID NO: 56]
acgtgcatag ccaccata

SEQUENCE 4 [SEQ ID NO: 57]
acaagaaacg tacatccc

Gene #20
E.coli dipeptide permease

Sequence 1 [SEQ ID NO:58]

1 ACAACCCTNC AGTGCTTGGC CAATTAGGTA GAGAATTTNA CCTAGGTAAN

51 TTAATGCGAT AAAGCCCAAG TTTGTAAAAT GTCCNTTGTG CGCCAATTTG

101 TTCCTGTACN TANTGGGANC TATTTTAGGA TTCTTATCAG GGATATTTCC

151 CAAGGGTTTT GTTGACNCCT TAATCATGCG TGCGTGTGAT GTTATGTTGG

201 CAATTCCCCA AGTTATGTTG TAACGTTAGC ATTAATTTGC ATTGTTTGGA

251 ATGGGTGCCG AAAATATTAT CATGGCATTT ATTTTGACGC GTTGGGCATG

301 GTTCTGTCGT GTTATACGTA CAAGTGTTAT GCAGTACACT GCTTCTGACC

351 ATGTCAGATT TGCTAAAACA ATCGGTATGA ATGATATGAA AATTATTCAC

401 AAACATATTA TGCCGTTAAC ATTAGCAGAT ATTGCTATCA TCTCTAGTAG

451 TTCGATGTGT TCAATGATCT TGCAAATATC TGGCTTTTCA TTTTTAGGAT

501 TAGGTGTCAA AGCGCCTACT GCAGAGTGGG GCATGATGCT TAACGAAGCT

551 AGAAAAGTGA TGTTTACACA TCCTGAAATG ATGTTTGNGC CAGGTATTGC

601 CATAGGGATT ATAGTGATGG CATTTAACTT CTTATCCGAT GCTTTACAAA

651 ATTGNTATTG GATCCCCCGC ATCTCTTTCT TAAAGATAAA CTTCCGCNCC

701 TTGTGAAAAA AGGGAGTGGN GCAATCATGA CATTGTTAAC AAGCTAAGCA

751 TTTGGCGATT ACAGATACCT GGACAGATCA ACCACCGTGA GTGATGTGAN

801 TTTNNCAATT AACTAAGGGG TGAAACTCTA GGCNTTATTG GGGAAAGTGG

851 TAGCGGT

SEQUENCE 2 [SEQ ID NO:88]

1 MGAENIIMAF ILTRWAWFCR VIRTSVMQYT ASDHVRFAKT IGMNDMKIIH

51 KHIMPLTLAD IAIISSSSMC SMILQISGFS FLGLGVKAPT AEWGMMLNEA

101 RKVMFTHPEM MFXPGIAIGI IVMAFNFLSD ALQNXYWIPR ISFLKINFRX

151 L*

SEQUENCE 3 [SEQ ID NO: 59]
atattatcat ggcattta

SEQUENCE 4 [SEQ ID NO: 60]
atctttaaga aagagatg

Gene #21
S.carnosus pts mannitol permease

SEQUENCE 1 [SEQ ID NO: 61]

1 GAATTCTTGC ACATGTTGCT CGGTGTCTTC CTTGCTGCAC TTGTATCATT

51 CGTTGTAGCT GCTTTAATTA TGAAGTTCAC TAGAGAACCA AAGCAGGATT

101 TAGAAGCTGC GACAGCTCAA ATGGAAAATA CTAAAGGGAA AAAATCAAGC

151 GTTGCTTCTA AGTTAGTATC TTCTGATAAA AATGTTAATA CAGAAGAAAA

201 TGCTAGTGGT AATGTTAGTG AAACATCTTC ATCAGATGAT GATCCTGAAG

251 CGCTATTGGA TAATTACAAC ACTGAAGATG TTGATGCACA CAATTACAAT

301 AATATAAATC ATGTTATTTT TGGCTGCGAT GCGGGTATGG GTTCTTNGGT

351 GCAAATGGGG TGCAAGCATT GTTACNGTNA TTAAATTTTA AAAAGGCGGC

401 AATTAATGAT ATTACAAGGG TACAAATTAC TGCGAATTAA TCAAATTGCC

451 AAAAGATGCT CCAATTANGN TATCAACTCC AGAAAAACTA CTTGATCCGG

501 GCTATTAACA AACACAATGC CATCCATATT CNAAGGGGNT TAATTTCCTA

551 ATCACCAAGA TATGNAGGAC TTTTAATTAT CTTAAAAAGG TGG

SEQUENCE 2 [SEQ ID NO:89]

```
       1  MIFGKGTAKA TSYGAGIIHF LGGIHEIYFP YVLMRPLLFI AVILGGMTGV

      51  ATYQATGFGF KSPASPGSFI VYCLNAPRGE FLHMLLGVFL AALVSFVVAA

     101  LIMKFTREPK QDLEAATAQM ENTKGKKSSV ASKLVSSDKN VNTEENASGN

     151  VSETSSSDDD PEALLDNYNT EDVDAHNYNN INHVIFGCDA GMGSSAMGAS

     201  MLRNKFKKAG INDITGYKYC D*
```

SEQUENCE 3 [SEQ ID NO: 62]
tgcacatgtt gctcggtg

SEQUENCE 4 [SEQ ID NO: 63]
GTGGTAATGT TAGTGAAAC

Gene #22
Mycobacterium phosphate sensor PhoR

SEQUENCE 1 [SEQ ID NO: 64]

```
       1  GGCACGAGCG AGTTCATTAG CTATATATAA GCCTAATCCA GAACCACCCG

      51  TTTTTGTATT ACGAGAGTTT TCTACTCTGA ATGTACGTTC GAATATACGT

     101  TCTTGTAGTT CTGGTATAAT GCCAATACCT CNATCGCTAA TAGCAATGTC

     151  GATAGTATCT TGATCTTTGT TTTCACTAAT ATTAATATCA ATGCGACTAC

     201  CAACATTTGA AAATTTTAGC GCATTATCAA GTAAGTTTGT TAAAATACGC

     251  TCAAGTGGCG TTCGATATTG ATAAAATGCA TCAATTTCGC TACAGAAATT

     301  CACTTCTAAT GTGCGGTTTT CATGTTTGAT ACGTTGCTCC ATATGGTTGC

     351  AATATTGATA CAAGTAATTG GTCTAGTTGT ATTAATTCTG GGGGATATGT

     401  TTTACCTGTA TTTAAAGTTG ATAAT
```

SEQUENCE 3 [SEQ ID NO: 65]
tataagcctaatccagaacc

SEQUENCE 4 [SEQ ID NO: 66]
aacgtatcaaacatgaaaac

Gene #23
UNKNOWN

SEQUENCE 1 [SEQ ID NO: 67]

```
       1  GTACGAGCTC GTGCCGGCAC GAGCGATTGG TGCAGTGAGT TATGTTTTAG

      51  AACAATTAGA TGCACCAGTA TATGGATCTA AATTGACAAT AGCGTTAATT
```

101 AAAGAAAATA TGAAAGCCCG TAATATTGAT AAAAAAGTTC GCTACTACAC

151 AGTTAACAAT GATTCAATTA TGAGATTCAA AAACGTGAAT ATTAGTTTCT

201 TTAATACGAC ACACAGTATT CCTGATAGTT TAGGTGTCTG TATTCACCCT

251 TCATATGGTG CCATTGTGTA TACAGGTGAA TTTAAGTTTG ACCAAAGTTT

301 ACATGGACAT TATGCACCAG ATATTAAACG TATGGCAGAG ATTGGTGAAG

351 AAGGCGTATT TGTCTTAATC AGTGATTCTA CTGAGGCAGA GAAACCTGGA

401 TATAATACTC CCGGAAAATG TAATTGAACA TCATATGTAT GATGCCTTTG

451 CCAAAGTGCG AGGTC

SEQUENCE 3 [SEQ ID NO: 68]
tttagaacaattagatgcacc

SEQUENCE 4 [SEQ ID NO: 69]
tccgggagtattatatccag

Gene #24
Anabaena nitrogen fixation gene

SEQUENCE 1 [SEQ ID NO: 70]

1 GGCCCAAACC CATCCAAGTC CTTTTTAATT GACTTATTTA CATTATTTCT

51 TTAATTTGGA TTAACAAATT TTTTTCTATT TGANCCCTTT AATGTTNACT

101 CCCCGTATCT AACAAGCAAG TGATCATACT TCATTATTTT AGCAACTCCT

151 TAATTTCCTC ATAAATGATG ATAAATATTT CTTTAAACCT TGCTATATCT

201 TCTTTAGTTG TAGTAGCCCC AAATGATAAT CTTATACTAC CTTCAATAGA

251 TTTGTCTGAT AATCCCATTG CAGCCAATAC TTCATTTAAT TTATTACGTT

301 TAGATGAACA AGCACTCGTC GTAGATATCA TAATGTCATA TTTTGAAAAA

351 GCATTAACTA ATACTTCACC TTTTACGCCA GGAAAACTAA GATTTAAAAC

401 GAATGGTGAA CCTGAAGTTG AAGAATTAAT ATAAACTCCA TGATATTTAT

451 TTAAAAATTG ACGGACGTCA TTATTTAACT CAGTAACAAA TGCATTCAAT

501 GCTTCAAAGT TTTCATTAGC TCGTGCC

SEQUENCE 3 [SEQ ID NO: 71]
ttttagcaactccttaatttcctc

SEQUENCE 4 [SEQ ID NO: 72]
gcacgagctaatgaaaactttg

Gene #25
UNKNOWN

SEQUENCE 1 [SEQ ID NO: 73]

```
      1  GACAACTTGC TAAAGCACGT GATGAAAAAG TAAGTGAATA TGGAATTGAA

     51  CAAGCTGATG GTACATTAAT TCAATATGAT AGTGAAGCCA AGATATATGA

    101  ACATTTTAAT GTGAATTTTA TACCACCTGC TATGCGAGAA GATGGTAGCG

    151  AATTTGATAA AGATCTAAGT AATATCATTA CATTAGATGA TATTAATGGT

    201  GATATTCATA TGCATACAAC GTATAGTGAT GGTGCGTTTT CTATTCGAGA

    251  CATGGTAGAA GCAAATATCG CAAAAGGTTA TAAATTCATG GTAATTACTG

    301  ATCATTCACA AAGTTTACGT GTTGCTAATG GCTTACAAGT GGAAAGACTT

    351  TTTANGACAA AAACGAAGGA AATTAAGGCT TTAGATAAAG AATATAGTGA

    401  AATTGGATAT TTATTCAGGT ACAAGAAATG GATATATTAA CCTGATGGCT

    451  CGCTGGATTA TGATGATGAA ATTTNAGCAC AACTTGGATA TGTNATTGGA

    501  GCTATTCAAC AAAGCTTNAN CCAATCAGAA GAACAAATNA TGGAACGGAT

    551  TAGCTAATGC ATGTCGCAAT CCATACGTGC GACATATAGC GCATCCAACA

    601  GGGCGTATTA TAGGTAGAAG AGATGGTTAT AAACCGAATA TTGAACAATT

    651  AATGGCATTA GCTGAAGAAA CGAATACAGT ATTAGAAATT AATGCCAATC

    701  CACATCGACT GGATCTTGAA CGCTGAAATC GNTCGNNAAT ATCCAAATGT

    751  GAAATTAACT NTTAACACTG ATGGGCATCA TNCAAATCAA TTNGATTTTN

    801  TGGAATTATG G
```

SEQUENCE 3 [SEQ ID NO: 74]
acgtgatgaaaaagtaagtg

SEQUENCE 4 [SEQ ID NO: 75]
tcttgtacctgaataaatatcc

Gene #26
periplasmic binding protein

SEQUENCE 1 [SEQ ID NO: 76]

```
      1  AGATCGTTCG CTAATTGACA ATTGATTAAA TCCCCTATTA CAAAATTGGA

     51  TATTACCTGT TATATCTAAA AATCCACAAA TTGCTTTAGC AAGTGTTGAT

    101  NTGNCGGCAC CATTGTGACC AACTATACTA AGCATTTCTC TTCTATAAAC
```

151 ATTTAATTGA ACATTATTAA GTACACTATT ACTATAGTCA CTATATTGAA

201 CACATACCTC ATTTAATTCT AATAGCGGCN CAGATGTGTA CTTATTATCA

251 TTATGTGCAG ATGTNTCATC TATCCATTTN NNCACTTTAA NTTTAACATG

301 TTCACTCATA CAAACGACAC GTAANTTCGC TAAGTTATCA ATGGATTCGA

351 CATCTACTTC TGNATATTNA AGCGCTGNAC AGTATAATGG NACACGTATG

401 CCTGCTTCTT TAAGCTTAGA TGATTTTAGC AAATCACTAG GCGTTGTATT

451 AGCGATGATT TTTCCATCTT TAAAAAGAAG ANCTCTATCA AACGTATCAT

501 CTAATGANTC TTCTAATCGA TGTTCGACAA TAATCATCGT TGACTTTGTT

551 TCTTCATGAA TATTGTNTAA CAATCTCAGC GTTTCATGTC CTGTCGCAGG

601 ATCTAAATTG GCCAGCGGCT CATCCAATAT TAAAATAGGC GTNCGATGGA

651 TTAATATACC ACCTAATGAA ACGCTCGTGC C

SEQUENCE 2 [SEQ ID NO:90]

1 GTSVSLGGIL IHRTPILILD EPLANLDPAT GHETLRLLXN IHEETKSTMI

51 IVEHRLEXSL DDTFDRXLLF KDGKIIANTT PSDLLKSSKL KEAGIRVPLY

101 CXALXYXEVD VESIDNLAXL RVVCMSEHVK XKVXKWIDXT SAHNDNKYTS

151 XPLLELNEVC VQYSDYSNSV LNNVQLNVYR REMLSIVGHN GAXXSTLAKA

201 ICGFLDITGN IQFCNRGFNQ LSISERS

SEQUENCE 3 [SEQ ID NO: 77]
aattgacaattgattaaatcccc

SEQUENCE 4 [SEQ ID NO: 78]
gccaatttagatcctgcgac

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: Burnham, Martin
Hodgson, John

(ii) TITLE OF THE INVENTION: Novel Compounds

(iii) NUMBER OF SEQUENCES: 91

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: SmithKline Beecham Corporation
(B) STREET: 709 Swedeland Road
(C) CITY: King of Prussia
(D) STATE: PA
(E) COUNTRY: USA
(F) ZIP: 19406-0939

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE: 25-FEB-1997
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 9604045.6
(B) FILING DATE: 26-FEB-1996

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Gimmi, Edward R

(B) REGISTRATION NUMBER: 38,891

(C) REFERENCE/DOCKET NUMBER: GM50007

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 610-270-4478

(B) TELEFAX: 610-270-5090

(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2111 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CTAGGAGTAG TATTTGGTTC ATGATTGCCT AATTCAATCA CATCTTTACT TTGCTCTAAG 60
TGCAAATCAC GCAATTGACC ATNTGGATCT CGTCTATCAT AGTCATAAAT ACGGTATGTC 120
GTATCGGATG ATTGTTGTGT CTCTAAAATT AAAATACCCG AACCAATGGC ATGGACAGTG 180
CCAGCAGGAA CATAATAAAA GTCACCGGGC TTAACAGGTA TACGTTTGAA AAGACTGCCA 240
AATTCATGAT TATCAATCAT GTCGATTAAC GCCTGTTTAT TATGTGCATG GACGCCATAA 300
TATAATTTCA GCACCTGGGC TGCATCTAAA TATACCAACA TTCTGTTTTA CCTAGTTCGC 360
CTTCGTGTTT TAAAGCGTAG TCATCATCTG GATGAACTTG AACAGATAAT TTATCATTGG 420
CATCTAATAC TTTAGTTAGC AGAGGGAAAC TATCTCGTGA ATCATTATCG AATAATTCAC 480
GATGTTGTGA CCAAAGTTGA TCTAGGGTCA TATCCTTGTA TGGACCATTG ATAATTGTAT 540
TAGGACCATT TGGATGTGCA GAAATTGCCC AGCATTCACC AGTTGTTTCA TTAGGGATAT 600
CATAGTTAAA TGCTTTTAAT GCATGACCGC CCCAAATTCT GTCTTTAAAA ACGGGTTGTA 660
AAAATAATGC CATAGTTAAA ACTCCTCTAT ATTTTCATTA ATAAGTTATA AATTTCTGTA 720
GTACTGTTGG CATTAATTAG TGATTGGCGT GTCTCATCAT TCATTAACGC TTTAGATAAG 780
CGCTGAAGTA TTTTTAAATG TGTATCCTGA CTGTTGTTTG GTACGGCAAT TAAGAATATC 840
AATTGAGGTA GACTACCATC TAGACTGTCC CATTTAACAC CATGATTATT TTTCATAACA 900
GCTACAATCG GTTGTTTTAC AACATCAGAC TTTGCATGTG GAATGGCCAC GTTCATGCCA 960

ATAGCTGTCG TAGACTCCAT TTCACGTTCT AGTATTGCAT TTTTTAAATG CGATGTGTGC 1020
TCTACATAAC GGCAAATTTT AAGTTTATGA ATCAACATAT CAATTGCTTC GTTTCGAGAC 1080
ATGTCGTGAT CAGTAATTAT CATAGTTTGT TGATCAAAAA CATGAGAAGG TTTATTGAGA 1140
TGTGAATGTT TCGCTCGTGC CATCNACATT GTCAACCTCT GTATCATGTT GTGTAATATC 1200
TGTATCATGA AGTTGCGTGT GTTGCGCTGG TGCATCTACT GCTATAACTG GTGTATTGCG 1260
TNTTAATAAT AGTACAGTAG GCATTGTGAC AAGACTACCT ACTATCNCTC CAAAGATAAA 1320
CCATAATACA TGATCAATAC CACCTAATAC AGCCACGATT GGACCTCCAT GTGCGACTCT 1380
ATCGCCGACA CCACCAATGN CTGCAATGAC TGATGCAATC ATTGCACCAA TGATGTTTGC 1440
AGGTATAATG CGCAATGGAT CTTGGGCTGC GAAAGGAATA GCACCTTCAG TAATNCCAAA 1500
TAGTCCCATA GTGAAGGNAG CCTTACCCAT TTCTCTTTCG GAATGATTGA ATTTATACTT 1560
NTGAACANAC GTTGCTAAAC CTAAACCGAT TGGTGGTGTA CATACANCAA CTGCGACCAT 1620
ACCCATAACG GCGTAATTAC CTTCAGCAAT AAGTGCTGAG CCAAATAAAA ATGCTACCTT 1680
GTTTAATTGG ACCGCCCATA TCGAAGGCGA TCATCGCACC TATAATCATC GACAAGTATA 1740
ATAATATTAG CACCTTGCAT ACTTTTTAAC CAGGGTTGTT AGGAATGCCG CAAAAATATT 1800
AGAAATCGTG CACCGATTAA AAATATAAAT ATCAATCCTA ACAACGACCG ATGAAATAAT 1860
GGGAATAATA ATGATAGGCA TAATTGGTGC CATTGCTTTT GGAACTTTAA TATCTTTAAT 1920
CCACTTTGCG ATATAACCTG CTAAGAAACC AGCAACAATA CCACCTAAAA ATCCTGCGCC 1980
TGCATCACTG CCATAAAAAC TACCGTCAGC AGCGATAGCG CCGCCAATCA TACCAGGAAC 2040
AAGACCGGGC TTGTCAGCGA TACTAACAGC GATATATCCA GCTCGTGCCG AATTCGGCAC 2100
GAGCTCGTGC C 2111

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ACCCTCTGTA TCATGTTG 18

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GTGCGATGAT CGCCTTGG 18

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 809 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

CGGCTCTTCG TAATATTGAT AATGTGCAAT ATTTNAAGAA TAATCAATTT ATTGAAGAAG 60
AAACCGTAGT GACCGTGAGC GAATATCGAA NCGGCTATTG ATAGAATACG TACTGAAATG 120
GACCCGAATG AATATCGAAG NCGATATAAA TGGTAGACCT AAACATATTT ACAGTATTTA 180
TCGGNAAATG ATGAAGCAGA AAAAACAATT TGATCAAATT TTTGATTTGT TGGCGATACG 240
TGTTATTGTC AATTCTATTA ATGATTGTTA TGCGATACTT GGGTTGGTGC ATACGTTATG 300
GAAACCGATG CCAGGACGTT TTAAAGATTA TATTGCAATG CCTAAACAAA ATTTGTATCA 360
GTCATTGCAT ACTACAGTAG TAGGTCCAAA TGGAGACCCG CTCGAAATCC AAATACGAAC 420
GTTTGATATG CACGAAATTG CTGAGCATGG TGTTGCAGCA CACTGGGCTT ACAAAGAAGG 480
TAAAAAAGTA AGTGAAAAAG ATCAAACTTA TCAAAATAAG TTAAATTGGT TAAAAGAATT 540
AGCTGAAGCG GATCATACAT CGTCTGACGC TCAAGAATTT ATGGAAACCT TATAATATGA 600
CTTACAGAGT GACAAAGTAT ACGCATTTAC CCCAGGGAGT GATGTTATTG AGTNGGCATA 660
TGGTGCTGTG CCGATTGGAT TTTGGCTTAT GCGAATCACA GGGAANGTAG GTAATAAGAT 720
GATTGGCGCC CAGGTGGAAT GGCAAAATTG TACCANATTG ACTTATNTTT TCACAAAACA 780

GGCGGATATT GTTGGAAATA CCGTTCTAG 809

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AGATACGTAC TGAAATGG 18

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CCTGTGATTC GCATAAGC 18

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1090 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GTGATGTGGC TAAACGCTTA AATGCAAATA TATATGTGTC TGGCGAAGGT GAAGATGCAT 60
TAGGGTATAA AAATATGCCA TCAAAAACAC AATTTGTTAA ACATGGAGAT ATCATTCAAG 120
TAGGCAATGT TAAATTAGAA GTTCTGCATA CTCCAGGACA CACGCCTGAA AGTATTAGCT 180
TTTTACTCAC TGATTTAGGT GGTGGNTCAN GTGTTCCGAT GGGATTATTT AGTGGTGACT 240
TTATTTNTGN TGGTGATATA GGTAGACCTG ATTTATTAGA AAAATCTTGT TCAAATAAAG 300
GGTTCGGCAC GAAATTAGCG CGAAACAAAT GTATGAGTCC GATCAAAATA TTAAAAATTT 360
ACCAGACTAT GTTCAAATCT GGCCGGGTCA TGGTGCTGGA AGCCCTTGTG GTAAAGCATT 420
AGGTGCCATA CCTATATCTA CAATAGGTTA TGAGAAAATT AATAACTGGG CATTTAATGA 480
AATTGATGAG ACTAAATTTA TTGNNTCATT AACATCAAAT CAACCAGCAC CACCNCATCA 540
TTGTGCACAA ATGAAACAAG TTANTCAGTG TGGCATGAAT TTATNTCAAT CATATGATGT 600
TTATCCNAGC TTAGATNATA AGAGAGTAGC ATTTGATCTT CGCGTAGCAA AGAGGGCTTT 660
CACGGGTGGC CACACAAAAG GAACAATCAA TATACCATAC AACAAAAACT TTATTANTCA 720
ANTTGGGTGG GTACTTAGAT TNTGAAAAAG ATATAGATTT AATTGGAGAT AAATCTACTG 780
TTGAGAAAAG CGAAACACAC TTTACAATTA ATTGGGTTTG ATAAGGTAGC AGGCTATCGT 840
NTGCCAAAAT CAGGCATTTC ACCCCAGTCC GNTCATAGCG CTGATATGAC AGGTAAAGAA 900
GAACATGTAT TAGACGTACG TAATGATGAA GAGTGGAATA ATGGACACTT AGNTCAAGCA 960
GTTAATATTC CACATGGTAA ATTATTAAAT GAAAATATTC CTTTTAATAA AGAGGATAAA 1020
ATATATGTAC ATTGTCAGTC AGGTGTTAGA AGNTCAATTG CAGTGGGGTA TATTGGGAAA 1080
GCAAAGGCTT 1090

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TTCGGGTGTT TTACCTTC 18

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TGCAGCAAGC CTTTTCTC 18

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2247 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AGCAGAATCT TTTTTAGCAT GATCTGTCAT AATGATCATA CGCTCTGGAT TTAAATCAGC 60
TAAATGTTCA GTGTCTAATT GTAAGTAAGG TCCTTTCAAA TATTTACTTA AACCTTGTGT 120
TACATCGTCA CTTAATGCAT TTTTAAATCC TAGNTCGTTT AAAAATTGTC CAACATATGA 180
ATAGTGTGGA TGTGCTAATA AACCAGCTTT AGCAACTACT GCTGGAAGCA CTTTGTGATT 240
TCTATCAAAT TTAATTTCAT CTTTATACTT ATTGATTAAT TTATCATGCT CAGCAAGACG 300
TTTNNCGCCT TCTTTNTCTT TATTTAAAGC TTTAGCAATT GTTGTTGAAC GAATTAATAT 360
TGTGGGTGTA GTCTCCATCA AAACTCTTTA ATGATAATGT GGTGCAATGT GGGCTAATTC 420
TTTATTAATA CCCTTATGTC TACTGCTATC AGNGATAATT AATCCCGGNT TTAATTTACT 480
AATNTCTCTT AAGTTNGCTT GTTACGTGTA CCTACAGAAG TATTACCCCC AATTTTTCTC 540

TTACTGGGTT ATGATACGTT TTTTCTTACC ATCATCAGCA ATACCAACTT GGTNTAACGG 600
CTATATGCTG NTAATGCAAC CTTGCAAATG AGTACTCTAA TACAACGATA CGTTGTGCAT 660
CTTTAGGTAC TTTTACTGTA CCATTTTCAT CTTTTACCCG AAATAGTATC TTTAGTTGAT 720
GATTCTTCTT TTACTTGAAT TATCCGTATT ACCACAAGCT GCAACTAAAA GTAAGGCAAC 780
TATTAATCCC AATATACTAA AAGTTTTTAG ACCTCTCATC NGTCCCACTC CTTAATATGT 840
ATANCTTCAT TTATTATTTT ATTGATAACA ATTATCATTG TCAAGTAGCG TTCAATCTTT 900
TTTATATTTC TAAAATGTAT GACTATATAT TTCCTCTAAT AATTATGACT ACAATTAGCA 960
CATTTCCTTA GACAAAATAC TGATAATGTA TCATTGCTAT ATCATCTTTG CATTAATACA 1020
ATTGACACCA CTTAGCATGA CCGNTATCCC TGTAATTCAG CTGATATTAT CTGTTGCAAT 1080
TTTATGTGAC GAACTGTTGC ACTTAATTTG ATAANTCAAC AANTACAANA NATCTAAGTT 1140
GAACAATTAT GATACAACCG TGCAAACGAT ATGTAGTATA ACTTGTCAAC TTAGAATTAT 1200
TGATAAATAT ATTAATATTG GTTTACCATA GCAGGAGATT TCACATCAAA ATTTTGAAGT 1260
AGCGTATCAA TCTTTGAATC ATCAATATAT ACCTTATGTA AATTTTTCAT ATACATCGAA 1320
TGAGAAAGTG CTTCATAATT TAATGAAAAA GATATATGAT CTCCAACTTG ATAGTGTCCT 1380
TGACCATTTA AATCAAGCAT TAAATGATCA CTCGAAGCGC CTAAAATATT GATATGCTGA 1440
TCCATAGGTG AAATATTATC GACTTGTGTA TCTNAAATAA CCAATATCTA CAATAGCTTG 1500
TAAGAATGAT TCATGCGTGT GTGTATTAAC TCGAGGTTTA ATTTCTAAAA TCTCAGCCTC 1560
CAATGTAATC GCATCTTGAT ATAACATAGC GAATCGCTTG ATTTGCGTTG TTTCAACAAC 1620
TCTAAACAAC GTNTCANCTA TTCGGAANTC AATTTATTTT TACCCAAATC AATATATAAA 1680
AGGTGGGGGG NAACATGCTC CGAATTACCA CCCGGAAATA ATTTNCANTC GATATCCTAT 1740
TTCTCTTNCA ACAGCTGAGA CGAATCGATT AATCATAAAG ATATCANCAC CACTTGGCGC 1800
ATCAGATTTA AAACACATAA AATTGAATGC TAAACCTACA AAATGGATAT TTTNCAAGTG 1860
AATAATCTCT TTANTATAAT CTAAAACATC ATAAGTCAGA ACACCTTCAC GGACATCTTT 1920
CCAATCTACC ATTAATAAAA TCTTATGTTT TTTTCCTAAA ACTTCTGCTA CTTCATTTAT 1980
NTGATGTATG GTAGATAATT CTGTGTGGAT ACTCATATCA ACTTTCCTCT ATCATATCTG 2040
AAATCTCTTT TGNGGGAGGC GTACGCAATA ACGTATATGT TAAATCCTGA TCTGCAATAC 2100
TAATTATGTT ATCCAATCTG GATTCTGCAA CATGATTGAT ACCTAACGCT TTTAAGCTTN 2160
CTACAATGGT ACGGGCANCA GCTATACACT TAATTACTGG TGTGANTNGN ATATTTTTAC 2220
TTTGAAAACT NNGTGGAGGT ACTTGGG 2247

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TGTAAGTAAG GTCCTTTC 18

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TAATACTTCT GTAGGTAC 18

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1789 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GGCACGAGCG GCACGAGCGT GTTGTATCAA GATTTTGTAG GCAGTTTTAC AACGTCCGAT 60
TCAGCAAGTT ATGCACAAGA TTTTAAATCT GAGGAAAACG CTAAAAAGAT TGCTGAAACT 120
TTAAATCTTT TATATCAATT AACAGGCAAT CAAAACGGTG TGAAAGTTGT GAAAGAAGTT 180
GTGGATAGAA CTGACTTGTC ATCTGATAAA TCAGTTGATA GCGAAACAAT GTAACTATAC 240

TAAGTTATGA GCATTACGCT CATAGCTTTC TTAGAAAGTA GGTGTAGTTT TGGATGATAT 300
TCAGAAAATA AAAAAAGAGC TTTCTGAATT AGTTGAACGT GTTGATGATG TTGAAATACT 360
AGCAAACGAA ACAGCTGATC ATGTGCTTGA ACTTAGAGAG GAACATAAGC AACATCATAA 420
TGAACTAAGA GAATCTCATA AAGAACTTAA AGATAAGCAA GATAAAGTTG TAGATGAGAA 480
TTTAGAGCAA ACAAAGATAT TAAACAGAAT TGAAGAAAGA TATCANACGC AAGTAGNTGT 540
TGNGCAAAAA AATGAAGAAA AGACACTCGC CCAAAATAAA TGGCTCGTAG GTGCCATATG 600
GGCGCTTGTA ACAATTGTTA TGATTGCAGT CATTACTGCA TCAATTNCTG CGTTATTACC 660
TTAAGGGAGG TGGACATAAT GAGTTGGGCA AGATGGTTAT CATGTTATTT GTNTGGTCGT 720
AAATGTAAAT AATGTTTTTG GTCAGTGCAT CGGCACTGGC TTTTTATTTT GATTGAAAAG 780
AGGTACGTAC ATGGTATTAC ACAGCTCACA AGACAGGAAG CATACTCCAA GTGAAGTTGG 840
GAAGTGTTGT TAATACCAAG TAAGTAGGAT ATCTGANATG TATAATAGAG TAAAAATGAA 900
ATCTTTTTAT TATAGACACA TATAAAAAGT GTATAGTAAT ATATGTATGT ATAATTAAAT 960
GATAATCATT TCATAATTAT TGTATATAAC TAAATAACTA CTTAACANAA ATAATTATGC 1020
TTTAGAGNTG ACCANNATGA NNNANNCCAG CATTTACATT ACTTTTATTC ATTGCCCTNA 1080
CGTTGACNAC AAGTCCCANT TGTAAATGGT AGCGAGAAAA GCGNAGNAAT AAATGCGAAA 1140
GATTTGCGAA AAAAGTCTGA ATTCCAGGGN ACAGCTTTAG NCAATCTTAN NCANATCTAT 1200
TATTACNATG NNANAGCTAN AACTGAAAAT AAAGAGAGTC CNCGACCACA TTTTTACAGC 1260
ATACTATATT GTTTANAGGC TTTTTTACAG ATCATTCGTG GTATANCGAT TTATTAGTAG 1320
ATTNTGATTC NNAGGATATT GTTNATAAAA ATAAAGGGNA AANAGTAGAC TTGTATGGTG 1380
CTTATTATGG TTATCAATGT GCGGGTGGTA CACCACACAA AACAGCTTGT ATGTATGGTG 1440
GTGTAACGTT ACATGATAAT AATCGATTGA CCGAAGAGAA AAAAGTGCCG ATCAATTTAT 1500
GGCTAGACGG TAAACANAAT ACAGTACCTT TGGAAACGGT TAAAACGAAT AAGAAAAATG 1560
TAACTGTTCA GGAGTTGGAT CTTCAAGCAA GACGTTATTT ACAGGAAAAA TATAATTTAT 1620
ATAACTCTGA TGTTTTTGAT GGGAAGGTTC AGAGGGGATT AATCGTGTTT CATACTTCTA 1680
CAGAACCTTC GGTTAATTAC GATTAATTTG GTGCTCAAGG ACAGTATTCA NATACACTAT 1740
TAAGAATNTA TAGAGATAAT AAAACGATTA ACTCTGAAAA CNTGCGTAG 1789

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

ATCCCCTCTG AACCTTCC 18

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

AAATGGTAGC GAGAAAAG 18

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3797 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TCAAATGCAG TCAGGGAAGC AATAGGACGA TATGCATAAA GGAGATGGTA AAGTGGAACA 60
GTGACAGAAG GTAAAGACAC GCTTCAATCA TCGGAGNCAT CAATCAANCA CAAAATAGTA 120
AAACAATCAG GAACGCAAAA TGATAATCAA GTAAAGCAAG ATTCTGGAAC GACAAGGTTC 180
TAAACAGTCA CACCAAAATA ATGCGACTAA TAATACTGAA CGTCAAAATG ATCAGGTTCA 240
AAATACCCAT CATGCTGAAC GTAATGGATC ACAATCGACA ACGTCACAAT CGAATGATGT 300
TGATAAATCA CAACCATCCA TTCCGGCACA AAAGGTATTA CCCAATCATG ATAAAGCAGC 360
ACCAACTTCA ACTACACCCC CGTCTAATGA TAAAACTGCA CCTAAATCAA CAAAAGCACA 420

AGATGCAACC ACGGACAAAC ATCCAAATCA ACAAGATACA CATCAACCCG CGTGCCTCAA 480
ATCATAGATG CAAAGCAAGA TGATACTGTT CGCCAAAGTG AACAGAAACC ACAAGTTGGC 540
GATTTAAGTA AACATATCGA TGGTCAAAAT TCCCCAGAGA AACCGACAGA TAAAAATACT 600
GATAATAAAC AACTAATCAA AGATGCGCTT CAAGCGCCTA AAACACGTTC GACTACAAAT 660
GCAGCAGCAG ATGCTAAAAA GGTTCGACCA CTTAAAGCGA ATCAAGTACA ACCACTTAAC 720
AAATATCCAG TTGTTTTTGT ACATGGATTT TTAGGATTAG TAGGCGATAA TGCACCTGCT 780
TTATATCCAA ATTATTGGGG TGGAAATAAA TTTAAAGTTA TCGAGGGAAT TGAGAAAGCA 840
AGGCTATAAT GTACATCAAG CAAGTGTAAG TGCATTTGGT AGTAACTATG ATCGCGCTGT 900
AGAACTTTAT TATTACATTA AAGGTGGTCA CGAGCGTAGA TTATGGCGCA GCACATGCAG 960
CTAAATACGG ACATGAGCGC TATGGTAAGA CTTATAAAGG AATCATGCCT AATTGGGAAC 1020
CTGGTAAAAA GGTACATCTT GTAGGGCATA GTATGGGTGG TCAAACAATT CGTTTAATGG 1080
AAGAGTTTTT AAGAAATGGT AACAAAGAAG AAATTGCCTA TCATAAAGCG CATGGTGGAG 1140
AAATATCACC ATTATTCACT GGTGGTCATA ACAATATGGT TGCATCAATC ACAACATTAG 1200
CAACACCACA TAATGGTTCA CAAGCAGCTG ATAAGTTTGG AAATACAGAA GCTGTTAGAA 1260
AAATCATGTT CGCTTTAAAT CGATTTATGG GTAACAAGTA TTCCGAATAT CGATTTAGGA 1320
TTAACGCAAT GGGGCTTTAA ACAATTACCA AATGAGAGTT ACATTGACTA TATTAAAACG 1380
CGTTAGTAAA AGCAAAATTT GGACATCAGA CGATAATGCT GCCTATGATT TAACGTTAGA 1440
TGGCTCTGCA AAATTGAACA ACATGACAAG TATGAATCCT AATATTACGT ATACGACTTA 1500
TACAGGTGTG TCTTCACATA CTGGTCCATT AGGGCACGAA AATCCTGCCG AATTAGGCAC 1560
GAGACATTTT TCTTAATGGA TACAACGAGT AGAATTATTG GTCATGATGC AAGAGAAGAA 1620
TGGCGTAAAA ATGATGGTGT CGTACCAGTG ATTTCGTCGT TACATCCATC CAATCAACCA 1680
TTTATTAATG TTACGAATGA TGAACCTGCC ACACGCAGAG GTATCTGGCA AGTTAAACCA 1740
ATCATACAAG GATGGGATCA TGTCGATTTT ATCGGTGTGG ACTTCCTGGA TTTCAACACC 1800
GTAAGGTGCA GAACTTGCCA ACTTCTATAC AGGTATAATA AATGACTTGT TGCGTGTGGA 1860
AGCGNCTGAA AGTAAAGGAA CACAATTGAA AGCAAGTTAA ATTCATCTTC TGAATTTAAT 1920
AGGCTATGTA AATCGTGCTG TTATCATGGC ACATCAGATA TAAGTAGCAT CACAGTGTTG 1980
AATCTCAAAA TAGTAAAGTG AAATAAAGCG CCTGTCTCAT TAGCGAAAAC TAAAGGGACA 2040
GGCGTATCTG TTTATGAGCT TAATAAATTG TATGAATAAT ATGGTTGATC GAATAACTGT 2100
TTATCATTGA TGATAAATTT GAGTTTTTTA AAAATAATTG ATATATTACA CCATTGTTAT 2160
AGCGTTTAAA GAAATCAACC CAACTTTACG ATAAATAGTG ATTGCTTCGT CATTAGGTCT 2220
ACGATCAAAA TCATGCTCGT TTTTATTCAC GCGTTCAAAT GTTGAATGTG GAACATGATT 2280
CATGATATGT TCGCTTTCCT CAACGGGAAC ATCATAATCG CCATTACAAT GCGCAATGAA 2340
AACAGGTGGA AGTGTTTTAA GNTCATCTGG TGCAATATTA TATTTTGAAT CAGTATAATC 2400
ANCAATGTTA ATCATATTTA TCCATTTACC TGTGCCACGT GCATAAACGT AGAGTAAAAA 2460
ACGTGTGCGA TTTGATCTTG ANCAACCGGT GTTGGTGAAG TGAGTTGTCC AATCATTGTT 2520
TCGTTTATGC TTTGAGCTAT TTTTGCGTAA TACCTATTAG TTGTTTTAAA AGGGTTCAGT 2580
GTTGATGCGA CTATAACCAT AAAAATCAAT AACACCATCA ATATCTCTGT CTCGTGCAAT 2640

TAATAAGACT TAAATATGCA CCTGATGATC TGCCAAAGGT AAAAATAGGG CAATTAGAAT 2700
ATTGTGATTG AATCGCATCG AATGATGCGT AGACATCCTC AATAATGCAA TCGAGACTTA 2760
CTTCTGGTAA TAAACGATAA CTTAGTTGAA TTAAATCGTA ATGTTCCGTA AGGATATCGA 2820
TATACTGTGG GGATAAATCG TTAGCTTTAC CGAACATTAA TCCACCACCG TGGATGTAGA 2880
CAATAACGCC TTTTGTTGGT TGATTTTTTG CTTTAATAAT TGTGTAAGGT AATGCAAATG 2940
CATCTTTAGT AATTACTTTA TATTTAATTT CAGTCACGAT TTAATAGGCT CCTTAGGAAT 3000
CCGATATTGA TGTCATTATA ACACTGTCNT NAATTTCCAT GNAAAATAGT CTTAAGACGA 3060
TGAGTCATGA TAATTCTGTT CCAATTGACG TAAAGCGTCN CGGGTATGCT TCTTTAGACC 3120
TTCCCCATAA TCCATCATTT TAACAATATC TTTAAAAGCA GCATGTGGNA TGGCTAAATC 3180
TTCTAAATCT GCCATAGAAA ATTCAAGATT GATATCATGT GGTCGCTGTT CAGCAAGTTT 3240
ATGCACAAAG TCAGGTTCTG TGACCAAAGG CGAAGACATG CCGACCATAT CTGCATGTTG 3300
TAAAGCATCT AAAGCAGACT CTGGAGAATT AATCCCGCCA CTTGCAATTA AAGGGATACG 3360
ACCTGCTAAA TGTTCATAGA CAATTTGGTT AACTGGTCGA CCGAAATGAT CACCTGGTGT 3420
ACGAGACGTA TTTTGATAAA TATGTCGACC CCAGCTAGCG ATTGCTAAGT ATTGGATGTT 3480
TGAAACGTCC ATGACCCAAT CGATTAATTG GTTGAACTCG TCAATGGTAT ATCCTAAATC 3540
ACTGCCTCTG GTTTCTTCTG GCGTTGCTCG AAATCCTAAA ATAAAATTGT CAGGTGCTTC 3600
TTTATCAATC ACTTCTTGTA CCGCACGCAT AACTTCTAAA CATAATCTTG CACGATTTTT 3660
TAATGAGTCG GCACCGTAAT GGTCTGTACG TCTATTTGAA AAAGTTGAGA AAAATGTTTG 3720
AATCAGCAAA CGTTGTGCAA TCGAAATTTC CACACCATCA AAACCTGCTT TAATCGCGCG 3780
TGCATCGAGC TCGTGCC 3797

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GACTAATAAT ACTGAACG 18

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TCTGTCGGTT TCTCTGGG 18

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1422 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CAGGCGTTTC CTCNGGTACN TGTTGCNNGC CTTTAATTAC CGACNCTGCA ATANCCAAAC 60
CGACCAGGTC GGATAGGGNA TATGTACCTG TTTTAGGACG ACCAATCGCT TGCCCAGTTA 120
AAGCATCCAC ATCTACNATG CTTANCTTGT GTTGCTCGGC GCGATACAGA ATATCATTCA 180
TTGTGTGCGT GCCGACTCTA TTTGCGACAA AGCCAGGCAC ATCATTGACG ACAATGACAC 240
CTTTACCTAA TACATTGTGC GCGAAATTTT TTACATCTAA TATGATAGAT TCCTTCGTGT 300
GTGACGTAGG TATTAACTCC ACTAATTNCA TAATACGTGG TGGGTTAAAG AAATGTAGAC 360
CAAAGAATCG CTCTTGATCC TTCTCGTTAA ATGCTTGAGC AATCGCATTA ATTGGGATTA 420
CCTGATGTAT TTGTAGCAAA TAAAGCATCT TCTNTAGCAT GTTGTAGAAC TTGTTGCCAA 480
ACAGCATGCT TAATTTCAAT ATCTTCTTTG ACTGCTTCGA TATATAAATC AGNATCATCA 540
TTTACCAAGT CATCATCAAA ATTACCATAT GTTAAATGAC TCACTAGATT TAAGTCGAAT 600
AGTAGCGGCC GTTTCTTATC TGTAATTTTA TCGTAAGATT TTTTCGCAAT GAGATTTGGA 660
TCGTTTGTGT CCACTACAAT ATCTAATAGT TTTACTTTAA GTCCAGCATN CACAAAGAGT 720
GCTGCCAGTT GAGCGCCCAT CGTGCCTGCG CCAAGAACGG TTACTTTATT AATTGTCATA 780

GTGATTCCTC CAATTTAGGT GAGGATAAGA TAACCATTAA GATAATTGGA ATAACGNTGC 840
TATTTTATNA AATTAATTAA GTATCTTTGA CAAGACATCT CAGNCTCTTT ATTTTAAGGA 900
AAAAGCTTTA TGCTTAAAAT AAGTCTTTTT TAGTGAAATT AATGCATCTC ATATAATTAT 960
TTGCTATTTA TACGAAAGCA GAATCTCCAG TCAAAGCGCG TCCAATTACT AAGGCATTAA 1020
TTTCATGTGT ACCTTCGTAC GTGTAAATCG CTTCTGCATC AGAGAAGAAA CGTGCAATAT 1080
CATAATCGTC AGCTAGTATG CCATTACCAC CTGTAATACC GCGGCCCATA GCTACTGTCT 1140
CACGCAAACG TAAGGCATTC ATCATCTTCG CCGGTGAAGT TGCAACCTCG TCATATTCAC 1200
CATGTGCTTG CATATTAGCT AATTGAGCAC ATGTTGCCAT TGCTTGAGCT AAATTACCTT 1260
GCATCATTGC TAGCTTNTCT TGTATTAACT GATATTTACT AATTGGGTNT GCCGAATTGC 1320
TTACGCTCAA GTGACATAAT CTAATGTGGC ACGTAAAGCG CCAGCCATAC CACCTGTAGC 1380
CATATAAGCA ACGCCTGCTC TCCGGTGGAA TAAAGAATTT TG 1422

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

ATGTACCTGT TTTAGGAC 18

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GAGTCATTTA ACATATGG 18

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 811 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

ATACTTTGAT TTTAGATGAA GCTGATGAAA TGATGAATAT GGGATTCATC GATGATATGA 60
GATTTATTAT GGATAAAATT CCAGCAGTAC AACGTCAAAC AATGTTGTTC TCAGCTACAA 120
TGCCTAAAGC AATCCAAGCT TTAGTACAAC AATTTATGAA ATCACCAAAA ATCATTAAGA 180
CAATGAATAA TGAAATGTCT GATCCACAAA TCGAAGAATT CTATACAATT GTTAAAGAAT 240
TAGAGAAATT TGATACATTT ACAAATTTCC TAGATGTTCA TCAACCTGAA TTAGCAATCG 300
TATTCGGACG TACAAAACGT CGTGTTGATG AATTAACAAG TGCTTTGATT TCTAAAGGAT 360
ATAAAGCTGA AGGCTTACAT GGTGATATTA CACAAGCGAA ACGTTTAGAA GTATTAAAGA 420
AATTTAAAAA TGACCAAATT AATATTTTAG TCGCTACTGA TGTAGCAGCA AGAGGACTAG 480
ATATTTCTGG TGTGAGTCAT GTTTATAACT TTGATATACC TCAAGATACT GAAAGCTATA 540
CACACCGTAT TGGTCGTACG GGTCGGTGCT GGTAAAGAAG GTATCGCTTG TAACGTTTGG 600
TTAATCCAAT CGAAATGGAT TATATCAAGA CAAATTGAAG ATGCAAACGG GTAGAAAAAT 660
GAGTGACTCC GCCACCTCAT CGGTAAGAAG TACTTCCAAG CACGTGAGGA TGACATCAAA 720
GGAAAAGGTG GAAACTGGAT GTCTTTAAGA GTCAAGAATC ACGCTGGAAA CGCATTCTTC 780
AGAGGTGGGT AAATTGAATT TTACGATGTG G 811

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GATGAAGCTG ATGAAATG 18

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TATCTAGTCC TCTTGCTG 18

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 960 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TAATTCGCAA TAGGAGTGAT GAATATCATA AATTTTACCC TCCAAATGAA GCTAATGAAG 60
TCCTGGACCC GAGTAAGACG CATGTAGCCA AGCTAAAATA ATCCACTCTA CCTTATCTTT 120
AGTTAATAAT GTTACTAAAT GTTGTTCATA CGCTGCTTTT GAATCAAATT GTTTTGGTTC 180
ATTAATATAA ACAGGAATAT CGTGCTTGTT TGCTCTATCT ATACAAAACG CATTTTGATG 240

ATCCGTATAT AGCNCCGTAA CTTCAATATT TTCAAGTTTT CCTGATTCAA CATGCTCAAC 300
TATATTTTCA AAGTTACTTC CTGAACCTGA TGCAAAAATC GCAATTTTAA CCATTGTTAT 360
ACCCCCAACA ATTCAATTGC AGTTGACTCA TTTTTCACAA TATGACCAAT TTGATAAGCT 420
TCCACATTTT GTTCTGCTAA AATCTTCAAA GCGCGTCGAT GCATCTTTTT CATCAACGAT 480
AACCGTATAG CCAATACCCA TGTTAAAAAT GTTATACATT TCATTTGTGT CTATATTGCC 540
TTGTTGTTGT AACCAATCAA ATATTTTTGG CGTTGGAAAT GATGTAGTAT CAATTCTAGC 600
AGCATATCCG GCTGGCAATG CACGTGGAAT ATTTTCATAA AAACCTCCAC CAGTAATATG 660
ATTCATTGCC TTAATAGAAA CTTCTTTTTT TAAAGCAAGT ACAGGTNTGA CATATAATTT 720
AGTTGGCTCT AAAAAGACAT CTATAAATGG ACGATTATCG NAGGGTGATG CCAAATCAAT 780
GNCTGATTCA NTAATTAATN TGCGCACTAA ACTGTNTCCA TTNGANTGAA TGNCACTTGG 840
ACGCAAGTCC TATAACAACT TGGCCCTCTT NCAATTCTTG AACCATCTTA CAATAGNCAA 900
CCTTTTTCAA CTGCTCCAAC AGCAAATCCG GCTACATCAT ATTCACCTTC GTGATACATT 960

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

ATAAGCTTCC ACATTTTG 18

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GATAATCGTC CATTTATA 18

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 541 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GGCACGAGCG CTAAATAATT AATATTTAGT TTTTAAGTTA TTAATAACGT AGGGATATTA 60
ATTTTAAAAG AAGCAGACAA AATGGTGTTT GCTTCTTTTT TATGTCGTAT AAGTAATAAA 120
TAAAACAGTT TGATTTTAAA ATGAAAGCGT AAAAATGGTA AAATATCCCA AAATTGATTG 180
TGATATAATT ATAAGGAAAA TGAGCAATTT ATGAAAAAAG TTTACGNACA AATCGGAGAA 240
TTAAAACTAA ATAATTATCA AAACAACGTC AATATTTAGT TGAATACTCA GACTTTAGCC 300
CATGGCCAAG TGGGGAAGAC AGCATATATT AGTAAAGGTG AATGATTTGT TATTACTCAC 360
TCGAAAATAG AAAGACAAGA TTTTAACGAT TAAAATAAAC TATTTTACAA ATAAAGTAAA 420
ATTAATTTAT TANGCTAATA ATGCAAAAAA TTAAAAAGTA ATGGACAAAG AGATAATGAT 480
ATGGCTCAAG AGGTAATAAA ATAGAGGTGG ACGCACACTA AATGGGGAAG TTAATACAAG 540
G 541

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GCACGAGCGC TAAATTTG 18

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CTTCCCCATT TAGTGTGC 18

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2334 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CCACCCANCT GATTATAATG TTTTAGCANG AGCTAGACTT GGTTGGTTAC CATCATATCC 60
ACAATTTAAT AAAAATAGTT TGTTGTTTGC AGAAGAAGCT AAAGATGAAG GCATTGAGTC 120
GAATGAGGCA ATTTTAAAAC GAGCGATAAA TGGAAGTTAA GTCAAAACAA ACGCAATTTG 180
CGATAGAAGA TCCGGATTTG AAAAAGAATC ATCCGGAAAT CACTGTTTAT ATGGCGCTCA 240
AATCTAATCT CAAGTTCTGC AAAAGGTCAA GAATACTTTA TGAAGCATTT ACTTGGCACA 300
AAATCAGGGT TATTAGCTAC ACCAAATGAA GATGAAAAGC CAGAAGAAAT TACGTGGCGT 360
GAGGAAACAA CAGGGAAATT AGATTTAGTC GTTTCTTTAG ATTTCAGAAT GACAGCAACA 420

CCTTTATATT CTGACATTGT TTTGCCAGCA GCGACTTGGT ATGAGAAGCA TGATTTGTCA 480
TCTACAGATA TGCATCCATA TGTACATCCT TTTAATCCAG CTATTGATCC ATTATGGGAA 540
TCGCGTTCAG ACTGGGATAT TTATAAAACG TTGGCAAAAG CATTTTCAGA AATGGCAAAA 600
GACTATTTAC CTGGAACGTT TAAAGATGTT GTGACAACTC CACTTAGTCA TGATACAAAG 660
CAAGAAATTT CAACACCATA CGGCGTAGTG AAAGATTGGT CGAAGGGTGA AATTGAAGCG 720
GTACCTGGAC GTACAATGCC TAACTTTGCA ATTGTAGAAC GCGACTACAC TAAAATTTAC 780
GACAAATATG TCACGCTTGG TCCTGTACTT GAAAAAGGGA AAGTTGGAGC ACATGGTGTA 840
AGTTTCGGTG TCAGTGAACA ATATGAAGAA TTAAAAAGTA TGTTAGGTAC GTGGAGTGAT 900
ACAAATGATG ATTCTGTGAG AGCGAATCGT CCGCGTATTG ATACAGCACG TAATGTAGCA 960
GATGCAATAC TAAGTATTTC ATCTGCTACG AATGGTAAAT TATCACAAAA ATCATATGAA 1020
GATCTTGAAG AACAAACTGG AATGCCGTTA AAAGATATTT CTAGCGAACG TGCTGCTGAG 1080
AAAATTCGTT TTTAAATATA ACTTCACAAC CACGAGAAGT AATACCGACA GCAGTATTCC 1140
CAGGTTCAAA TAAACAAGGT CGACGATATT CACCATTTAC AACGAATATA GAACGTCTAG 1200
TACCTTTTAG AACATTAACA GGACGTCAAA GTTATTATGT GGATCACGAA GTTTTCCAAC 1260
AATTTGGGGA GAGCTTACCA GTATATAAAC CGACATTGCC GCCAATGGTA TTTGGGAATA 1320
GAGATAAGAA AATTAANGGT GGTACAGATG CTTTGGTACT GCGTTATTTA ACGCCTCATG 1380
GANAATGGAA TATACACTCA ATGTATCAAG ATAATAAGCA TATGTTGACA CTATTTAGAG 1440
GTGTCCACCG GTTTGGATAT CANATGAAGA TGCTGNAAAA CACGATATCC AAGATAATGA 1500
TTGGCTAGAA GTGTATANCC GTAATGGTGT TGTAACGGCA AGAGCAGTTA TTTCGCATCG 1560
TATGCCTAAA GGTACAATGT TTATGTATCA TGCACAAGAT AAACATATTC AAACGCCTGG 1620
GTCAGAAATT ACAGATACAC GTGGTGGTTC ACACAACGCG CCGACTAGAA TCCATTTGAA 1680
ACCAACACAA CTAGTCGGAG GATACGCACA AATTAGTTAT CACTTTAATT ATTATGGACC 1740
AATTGGGAAC CAAAGGGATT TATATGTAGC AGTTAGAAAG ATGAAGGAGG TTAATTGGCT 1800
TGAAGATTAA AGCGCAAGTT GCGATGGTAT TAAATTTAGA TAAATGCATA GGATGCCATA 1860
CGTGTAGTGT GACATGTAAA AACACTTGGA CAAATCGTCC AGGTGCTGAG TAACATGTGG 1920
TTCAATAACG TAGAAACGAA GCCAGGTGTA GGGTATCCGA AACGTTGGGA AGACCAAGAA 1980
CACTACAAAG GTGGTTGGGT ACTAAANTCG TAAAGGGAAA CTTGAATTAA AATCTGGAAG 2040
TAGAATTTCA CAAATTGCTT TAGGTAAAAT TTTTTATAAC CCAGATATNC CATTAATAAA 2100
AGATTATTAT GANCCATGGA NCTATAATTA TGAACATTTA ACAACTGCGA AATCAGGGAA 2160
GCATTCGCCA GTTGCTAGAG CGTATTCAGA AATTACAGGG GATAACATTG AAATTGAATG 2220
GGGACCTAAC TGGGAAGATG ACTTAGCAGG TGGTCATGTT ACAGGCCCAA AAGATCCTAA 2280
CATACACAAA ATAGAAGAAG AGATTAAATT CCAATTTGAC GAAACTTTTA TGAG 2334

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

ATTGATCCAT TATGGGAA 18

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CATATTGTTC ACTGACAC 18

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 638 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

AGTTATTGTA TTTAAAAATG TTTCATTTCA ATATCAAAGT GATGCATCCT TCACATTGAA 60

AGATGTTTCT TTTAATATAC CTAAAGGTCA GTGGACATCT ATTGTTGGTC ATAACGGTTC 120
TGGAAAATCT ACAATTGNCA AGTTAATGAT TGGCATAGAG AAAGTTAAAT CTGGAGAAAT 180
TTTTTATAAT AATCAAGCTA TAACTGATGA TAATTNTGAA AAGTTAAGAA AAGACATAGG 240
AATTGTATNT CAGAATCCGG ATAATCAATN TGTTGGNTCA ATTGTAAAAT ACGATGTGGC 300
ATTTGGACTC GAAAATCATG CGGNTCCACA TGACGAAATG CATAGAAGAG TCAGCGAAGC 360
ACTTAAACAA GTTGATATGT TAGAACGTGC AGATTATGAC CCTAATGCAT TATCGGGGGG 420
ACAGAAGCAG CGTGTGGCTA TAGCAAGTGT ATTAGCACTT AACCCTCTGT CATTATATAG 480
ATGAGGCGAC TCTATGTTAG GATCCCTGAT GCACGTCAAA TTTATGGGAT TTAGNGAGAA 540
AGTAANTCAG ACATTATATA CAATCATTCT ATACGCATGA TTTATCTGAG GCGATGAGNA 600
GATCAAGTAT CCGTATGATA AGGACTTNCT TTTAAGGC 638

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GTTTCATTTC AATATCAA 18

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

ATCTATATAA TGACAGAG

18

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1496 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GTTAATCAAG TATCGAAGCG GAACAATCAT ACTTTAATGT TGAAGATTTA TATNGCGAAC 60
AAGCGATGGT CCTAGTGCGT AATATTAATT TAGCACTGCG CGCACAATAT TTGTTNGNAT 120
CTNATGTCGA TTACTTTGTA TATNNTGGTG ATATTGTTTT AACTGACCNC ATTACAGGTC 180
GTNTGTTACC GGNAACTAAG TTGCAAGCTG GACTTCACCA NGCTATTGAA GCGAAAGAAG 240
GTATGGAGGT TTCAACAGAT AAAAGTGTTA TGCCAACCAA TTACCCTTCC AGAATTTATT 300
TAAACTTTTT GAATCAATTT TCAGGTATGA CAAGCTACAG GAAAATTAGG CGAATCAGAG 360
TTCTTTGATT TGTATTCANA AATAGTCGTA CAAGCACCCA ACTGATAAAG CGATTCAACG 420
TATCGATGAA CCAGATAAAG TGTTTCGTTC AGTTGATGAG AAAAACATCG CGATGATTCA 480
TTGATATAGT TGAACTTCAT GANNCGGGGC CGACCGGTTT TACCTCATAA CCGAGNACTG 540
CTGAAGCGGC TTGAATACTT TTCNGAAGTA TTATTCCAAA TGGATATTCC TAATAATTTA 600
CTCATTGCGC AAAATGTTCC AAAAGAAGCG CAGATGATAG CTGAAGCAGG CCAAATTGGT 660
TCCATGACTG TTGCGACTAG TATGGCAGGT CGAGGCACAG ATATTAAACT TGGTGAAGGT 720
GTCGAAGCAT TAGCTGGATT AGCTGTTATT ATTCATGAAC ATATGGAAAA TAGCCGTGTA 780
GACAGGCAAT TACGTGGTCG TTCTGGTAGA CAAGGGGATC CGGGATCATC TTGTATATAT 840
ATTTCACTAG ATGATTATTT AGNTAAGCGA TGGAGCGATA GTAATTTAGC GGAAAATAAT 900
CAATTATATT CANTAGATGC ACAACGATTA TCGCAAAGTA ATTTGTTTAA TCGNAAAGTT 960
AAGCAAATTG TAGTTAAAGC GCAGCGTATC TCGGAAAGAA CAAGGGGTTA AAGCTCGGTG 1020
AAATGGCTTA ATTGAATTTG NNAAAAAGCA TNAGTATTCA GCGAAGATCT TNGTATTTAC 1080
GANGGAACGC AAATCCGAGT TTTTAGAAAT TAGATTGATG CTGAGAATCC NAGATTTTTA 1140
ANGCGGTTAG CTTAAAGATT GTATTTGAAA TNGTTTGGGG NAATGANGGA AANGGTGCTA 1200
ACAAAATCGC GNGTTGGGCG AGTATATTTT ATCAAAAATT TAAGTTNCCA ATTTAATAAA 1260
GATGTGGCTT GTGTTAATTT TAAAGATAAG CAAGCAGNAG TGACATTTTT ATTAGAGCAA 1320
TTTGAAAAGC AATTAGCTTT GGANTCCGTA AAAACATGCA ANGNGCATAT TATTATAATA 1380

TTNCCGGCCA AAANGTCTTT NGGGAAAGCA ATTGATNCAA GTTGGGGTTA GGAACAAGTC 1440
GGCTTTTNAC AACAANTTAA NAGCAAGCGN TAATCAAACG ACAAAANTGG CAACCT 1496

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CCGCTAAATT ACTATCGC 18

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CTGAAGCGGC TTGAATAC 18

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 955 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

ATATAAATTA TTTAAGCGTA TGGTTTTACT TCGATTGCAC CCTTCATTTT CATCATTGAA 60
CACCATGCTT AATATAATCC ATATATTTGT GGCTCTAAAG NCTTTCCTCC CACCGTATAA 120
TGTCTGCTGC TTTTTCAGCT AACATTAAAA CAGGTGCGTG TATATTGCCA TTTGTCGTAC 180
GTGGCATAGC GGATGCATCA ACTACACGTA AATTTTCCAT ACCGTGGACT TTCATTGTTA 240
ACGGGTCAAC TACTGCCATT GGATNCTGAA GCAGGACCCA TTTTAGCACN ACAAGATGGG 300
TGTAATNCTG TTTCACCATC TCNACGGAAN NCAATCAAGN ATTTCTTCGT CTGTTTGCAC 360
TTCTGGGTCC TGGGTGAAAT TTCTCCACCA TTGAATGGAT CCATTGCTTT TTGAGATAAG 420
ATATTTCTTG CTACACGAAT TGCTTCTACC CATTCTNTTT TATCTTCTTC TGTTGATAAA 480
TAATTAAAGC GGATACTTGG TTTTTCGAAT GGATCTTTAG ATTTGATTGG CACGAGCTAC 540
CACGAGAGTT TGAATACATT GGTCCTACGT GAACTTGATA ACCATGTGCG ACCGCTGCCT 600
TTTGACCATC ATATCTTACA NCTATTGGTA AGAAATGGAA CATTAAGTTA GGATAATCAA 660
CTTCGTTATT TGAACGTACA AATCCGCCAC CTTCAAAATG GTTAGATGCT GCTGCACCTG 720
TACGTGTGAA AATCCAGTGG TAAACCAATT AAATGGCATG CGCCTTGATA TCTAAGCTTG 780
GCTGTAATGA TACAGGTTTC CTTACATTTA TGTTGAATGT ATACCTCTAA GTGATCTTCC 840
AAAGTTTTCA CCCACACCTG GTAAATGAAC ACGTGGCTCA ATGCCTTTTG ATTTTAGGAA 900
CTCTGAATCA CCGATACCAG ATAATTGTAG TAATTGTGGC GTTATTGAAT GCCCC 955

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GAAGCAGGAC CCATTTTA 18

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GATTTTCACA CGTACAGG 18

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 497 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GAATTCCTAC ATAATACTTT TGTTTACCTT GTGTCAGTTT ATACAACGGT GGCTGTGCAA 60
TATACACATA GCCTGCTTCA ATTAACGGTC TCATAAATCG ATAGAAGAAT GTTAATAACA 120
ATGTTCTAAT ATGCGCTCCA TCCACATCGG CATCAGTCAT AATGACGATT TTGTGATATC 180
TTGCTTTCGC TAGATCAAAG TCGCCACCGA TTCCTGTACC AAATGCTGTG ATCATTTGAC 240
GAATTTCATT GTTATTCAAA ATTCTATCTA ATCGTGCTTT NTCAACATTT AATATCTTAC 300
CTCGTAATGG TAAAATCGCC TGCGTTCTAG AGTCACGACA GATTTTGGTG GACCCCCNGC 360
AGAGTCCCCT TCGACTAAGA AAATCTCACA TTCTTCAGGA CTTTTACTAG AGCAATCGGC 420
TAATTTACTG GAAGACTGCT ACATCTACGC TGATTTACGA GGTGTTACTT CAGGGCTTTN 480
TCGAGACACG TGCANGT 497

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

CATAATACTT TTGTTTACC 19

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

AGTAACACCT CGTAAATC 18

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1443 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

CTANCNAANG GAANTTCAGC ATCCTTAAAA ATACCTATTT GACTGTAGAA ACCTTTTGNT 60
GCGTACAATA TCTAAACCTT GTCGTGCTGC TGGAACTGCA CCTGAACATT CAACAACAAC 120
ATCTGCACCG TAACCGTCTG TAATTCCATT GATATACGTT TTTAAGTCTG TGTGTTGTAA 180
ATTGACTACA TAATCCATGT GCAATGCTTC TGCTTTATCT AATCTGACTT NGTGGCANTG 240
TCCAATCCAG TTACCACAAC AGGTGCGCCT TTACTTTTCA ACACTTGTGC TACAAGTAAT 300
CCGATTGGCC CAGGTCCCAT TACAACTGCT ACATCGCCAG AGTTCACTTG AATCTTAGAA 360
ACGCCATGAT GTGCACATGC TAATGGTTCT TGTCATAGCT GCAGACTGAT ACGATACTTC 420
CGCTTCTGGA ATATGATNCA AACTTTCTTC ACGTGCAATG ACATAATTAG TAAATGCGCC 480
ATCAACTTGT GTTCCAATAC CTTTTCGATG GTTGCATAAA TGATAGTTTT TTGATTTACA 540
GGAATCACAC TCATTACANA CCATAGAATG TAGTTTCAGA AGTGACNCGG TCACCAACTT 600
TAAAATCNTT AACGTCTGCT CCCAACTTCA ACGATNTCAC CAGAAAATTC ATGACCTAAT 660
GTCACTGGAA AATTAACTTN ATAATGCCCT TCATAAGTAT GAAGGTCTGT GCCACAAATT 720
CCTGCATAAT GTACTTTAAT CTTTACTTTA TCATCTAGCG GTGTTGCAAC TTCTTTATCA 780
AGAAGTTCTA AGTTGCCATG TCCTTCTCTT GTTTTTACTA AAGCTTCCAC CACAAACACN 840
TCGANTTTTT ANTTGNAATA GACTNNATAG NTTNAAGATA AGATAGTTAN CGATATTNCC 900
ACCTTGATCA ATACTTGANA TTTCAGATGA ACCTTTTGNC ATTTGTACAT TCGTACCTTT 960
CGCCATATCT GTGAAAATGG GTGCTACGTC TGTTGCAATA TATAATGAAA TTGCAATCAT 1020
AATCGTACCC ACAATGACAG AATGAATAAT GTTTCCTCTT GCTGCACCAA CAATAAACGC 1080
GACAACAAAT GGTATAGTTG CTAAGTCACC AAAAGGTAGT ACTTGGTTTC CTGGTAAAAT 1140
AACGGCTAAT AAAACAGTGA TAGGTACTAA AATTAATGCT GTCGAAATAA CCGCTGGATG 1200
ACCTAATGCT ACAGCCGCAT CCAATCCAAT ATAAATTTCA CGTTCGCCAA AACGTTTATT 1260
TAGCCATGTT CTTGCAGACT CTGAAACTGG CATTAAACCT TCCATTAAGA TTTTTACCAT 1320
TCTAGGCATT AAGACCATTA CTGCAGCCAT TGACATTCCT AAATTAATGA TGTCTCCAGG 1380
TTTGTAACCT GCTAACACAC CAATACCTAA ACCTAAAATT AAGCCGACAA ATATAGACTC 1440
TCC 1443

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

GTTCTAAGTT GCCATGTC 18

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CCTAGAATGG TAAAAATC 18

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1642 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CCATTTAAAA GTATTGTAAA ATCATCCACN TTNTATAAAC CAACCACNTT AACNTTTTTG 60
ACATTTGTTA TCCGATGAGA TTAAAAGATA TCAATNAATA CAATTTTTAN AATTAATGTC 120
ACTATGTTTT CCGATAATAT NACCCAATCA TCGNAATGTT ACCCATTTAT AAAATGANAA 180
ATCNTTGACA TAGGTANAGG GAATGTATAT TGGTCNCGGA TCACTTAAAT TAAACCCANA 240
TCATGTCATC TGGTAATGTN TCAATGTTAA TTGCTCCTGA AGCGGCGTAN ACTTTAATCT 300
TCCATGTTAA ATGAGTAAAT TGATGCGTCA ACTCNAAAAT AGGTGTTTCT NCTGGNTGAA 360

TGTCATGACC GATTTTTTCA NTCATTTTAC GTCTANCATG CTCACTATCN AACATAGGAN 420
ATTGCCACAT ACCATACNAT AATTNTTCCC TACGCTTTTG CAACAGATAT TGACCTTGAT 480
TATTTCTAAT TAANAAGACG GATTGCTCAA TTACNTTTTT ACTTACATTT TTAGATTTAA 540
CAGGTAACTT TTCAAATGGA CCTTTATCAA ATGCCTCACA GTTTTCTTGN ACTGGACNAA 600
ATAAGCATAA TGGATTTTTT GGTGNACAAA TTAATGCCCC TAATTCCATC ATAGCTTGAT 660
TAAACGTTCC AGCTTCTGTA GTAACATACG GTAACAATTC TTGTTCGTAC GATTTCCTCG 720
TCGATTGTAA TTTAATATCT CGATAGTCAT CATTCAATCT AGACCATACG CGAAAAACAT 780
TTCCGTCTAC AGTTGCTAGT GGTACATTAT ATGCAATGCT CATTACTGCA GCTTGTGTGT 840
ATGGGCCAAC ACCTTTTAAC GCTTTAAATT GATCAGGATC TTTGGGAACT AAGCCTTCAT 900
ATTTATCANA AACTTCTTTA ATCGCCGTAT GAAAATTTCG AGCTCTACTA TAATATCCTA 960
AGCCTTCCCA ATACTTTAAC ACTTCATCTT CCGAAGCTTG ACTCAAAACT TCCACAGTTG 1020
GAAATCGGNC ACCAAAACGA TGATAATAGT CAATAACTGT TTTAACTTGT GTCTGTTGTA 1080
ACATGACCTC ACTTAACCAA ATATAGTACG GATTGGTCGT TTGTCGCCAT GGCATTTCTC 1140
TTTGATTTTC ATCAAACCAG TGTATCAAAT TTTCTTTAAA ACTAGACTGC TGATACATTT 1200
ATAAAACCCT TTCCTCACCA AAATTAATTG TCTTTACTCA TAATGTTTTT ATTGTACATT 1260
AAAATCATGG TTAGTATGTA AGTTAATTTA GTTATNTGCG AAATTGGATT ATAATAGTAT 1320
ATATAATATT ATGAAATGAG TGAACTGATA TGGACACTGC AACACATATC GCAATTGGGG 1380
TGGGCCTTAC AGCACTTGCA ACTCAAGATC CAGCAATGGC TTCTACGTTT GGTGCAACAG 1440
CTACAACCCT TATCGTTGGT TCATTAATTC CTGATGGGGA TANTGTNCTT AAATTANAGG 1500
ACANTGCAAC ATATATTTCG NATCATAGAG GNATNACGTC ATNCCATCCC CTCCCACAAN 1560
NNTATGNCCA GTCNCNTTTA CANTTTNTAT NTNTTCACGT CACTNTNGCT GGTANGCATC 1620
CCNCCTCACG TATGGCTTGT GG 1642

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

TCCTGAAGCG GCGTATAC 18

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

TATGAAGGCT TAGTTCCC 18

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 514 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GGGAAAAAAA GAAAACCTTC CAAAATACGG GAAATTGAAA TTAATTANCC GGAGAGACCA 60
NATAGGAAGT AATTGATAAT GGAAGTTTCC CCANAATTTA ACAAGCTAAA AGAGTTTGGG 120
TGCCTTTTAC AAGATAAGCA TGCCAATACA GTCATTTCAC GCACACTGTT GNCCACTATG 180
AGTTAAAGCT TGCTGAAGGT TATGAAACAC ATTTAGTGGG AATAAAAAAC AATAATAACG 240
AGGTCATTGC AGCTTGCTTA CTTACTGCTG TACCTGTTAT GAAAGTGTTC AAGTATTTTT 300
ATTCAAATCG CGGTCCAGTG ATCGATTATG AAAATCAAGA ACTCGTACAC TTTTTCTTTA 360
ATGAATTATC ANAATATGTT AAAAAACATC GTTGTCTATA CCTACATATC GATCCATATT 420
TACCATATCA ATACTTGAAT CATGATGGCG AGATTACAGG TAAGGCTGGT AATGATTGGT 480
TCTTTGATAA AATGAGTAAC TTAGGATTTG AACG 514

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

GAGGTCATTG CAGCTTGC 18

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

CAAATCCTAA GTTACTCATT 20

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 479 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

CGCACATAAC GTGCAGCATA TGCAGCTGAG CGGTCTACTT TTTGTAGGAT CCTTACCACT 60
GAAGCATCCG CCACCATGAC GTGCATAGCC ACCATACGTA TCAACAATGA TTTTACGTCC 120
TGTTAATCCT GCATCACCTT GAGGTCCACC GATTACAAAG CGTCCTGTAG GATTGATGTA 180
GAATTTAGTT TGTTCATTAA TCAAGTTTTC TGGAACAGTT GGATAAATGA CATGCGCTTT 240
GATGTCTTCT TGAATTTGTT CAAGTGTCAC ATCATCAGCA TGTTGTGTTG ATACGACAAT 300
CGTATCAATA CGTACTGGGT TATCATTTTC ATCATATTCA ACAGTGACCT GAACTTTACC 360
GTCTGGTCGT AAATAATTCA ACGTCTCGNG CCATCTTTTA CGCACATCAG ATTAAACGTT 420
TGGGGCAATT GGGTGTGATA AATTAAATTG CTAGAGGGAT GTACGTTTCT TGTTTCAAT 479

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

ACGTGCATAG CCACCATA 18

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

ACAAGAAACG TACATCCC 18

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 857 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

ACAACCCTNC AGTGCTTGGC CAATTAGGTA GAGAATTTNA CCTAGGTAAN TTAATGCGAT 60
AAAGCCCAAG TTTGTAAAAT GTCCNTTGTG CGCCAATTTG TTCCTGTACN TANTGGGANC 120
TATTTTAGGA TTCTTATCAG GGATATTTCC CAAGGGTTTT GTTGACNCCT TAATCATGCG 180
TGCGTGTGAT GTTATGTTGG CAATTCCCCA AGTTATGTTG TAACGTTAGC ATTAATTTGC 240
ATTGTTTGGA ATGGGTGCCG AAAATATTAT CATGGCATTT ATTTTGACGC GTTGGGCATG 300
GTTCTGTCGT GTTATACGTA CAAGTGTTAT GCAGTACACT GCTTCTGACC ATGTCAGATT 360
TGCTAAAACA ATCGGTATGA ATGATATGAA AATTATTCAC AAACATATTA TGCCGTTAAC 420
ATTAGCAGAT ATTGCTATCA TCTCTAGTAG TTCGATGTGT TCAATGATCT TGCAAATATC 480
TGGCTTTTCA TTTTTAGGAT TAGGTGTCAA AGCGCCTACT GCAGAGTGGG GCATGATGCT 540
TAACGAAGCT AGAAAAGTGA TGTTTACACA TCCTGAAATG ATGTTTGNGC CAGGTATTGC 600
CATAGGGATT ATAGTGATGG CATTTAACTT CTTATCCGAT GCTTTACAAA ATTGNTATTG 660
GATCCCCCGC ATCTCTTTCT TAAAGATAAA CTTCCGCNCC TTGTGAAAAA AGGGAGTGGN 720
GCAATCATGA CATTGTTAAC AAGCTAAGCA TTTGGCGATT ACAGATACCT GGACAGATCA 780
ACCACCGTGA GTGATGTGAN TTTNNCAATT AACTAAGGGG TGAAACTCTA GGCNTTATTG 840
GGGAAAGTGG TAGCGGT 857

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

ATATTATCAT GGCATTTA 18

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

ATCTTTAAGA AAGAGATG 18

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 593 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

GAATTCTTGC ACATGTTGCT CGGTGTCTTC CTTGCTGCAC TTGTATCATT CGTTGTAGCT 60
GCTTTAATTA TGAAGTTCAC TAGAGAACCA AAGCAGGATT TAGAAGCTGC GACAGCTCAA 120
ATGGAAAATA CTAAAGGGAA AAAATCAAGC GTTGCTTCTA AGTTAGTATC TTCTGATAAA 180
AATGTTAATA CAGAAGAAAA TGCTAGTGGT AATGTTAGTG AAACATCTTC ATCAGATGAT 240

GATCCTGAAG CGCTATTGGA TAATTACAAC ACTGAAGATG TTGATGCACA CAATTACAAT 300
AATATAAATC ATGTTATTTT TGGCTGCGAT GCGGGTATGG GTTCTTNGGT GCAAATGGGG 360
TGCAAGCATT GTTACNGTNA TTAAATTTTA AAAAGGCGGC AATTAATGAT ATTACAAGGG 420
TACAAATTAC TGCGAATTAA TCAAATTGCC AAAAGATGCT CCAATTANGN TATCAACTCC 480
AGAAAAACTA CTTGATCCGG GCTATTAACA AACACAATGC CATCCATATT CNAAGGGGNT 540
TAATTTCCTA ATCACCAAGA TATGNAGGAC TTTTAATTAT CTTAAAAAGG TGG 593

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

TGCACATGTT GCTCGGTG 18

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

GTGGTAATGT TAGTGAAAC 19

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 425 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GGCACGAGCG AGTTCATTAG CTATATATAA GCCTAATCCA GAACCACCCG TTTTTGTATT 60
ACGAGAGTTT TCTACTCTGA ATGTACGTTC GAATATACGT TCTTGTAGTT CTGGTATAAT 120
GCCAATACCT CNATCGCTAA TAGCAATGTC GATAGTATCT TGATCTTTGT TTTCACTAAT 180
ATTAATATCA ATGCGACTAC CAACATTTGA AAATTTTAGC GCATTATCAA GTAAGTTTGT 240
TAAAATACGC TCAAGTGGCG TTCGATATTG ATAAAATGCA TCAATTTCGC TACAGAAATT 300
CACTTCTAAT GTGCGGTTTT CATGTTTGAT ACGTTGCTCC ATATGGTTGC AATATTGATA 360
CAAGTAATTG GTCTAGTTGT ATTAATTCTG GGGGATATGT TTTACCTGTA TTTAAAGTTG 420
ATAAT 425

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

TATAAGCCTA ATCCAGAACC 20

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

AACGTATCAA ACATGAAAAC 20

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 465 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

GTACGAGCTC GTGCCGGCAC GAGCGATTGG TGCAGTGAGT TATGTTTTAG AACAATTAGA 60
TGCACCAGTA TATGGATCTA AATTGACAAT AGCGTTAATT AAAGAAAATA TGAAAGCCCG 120
TAATATTGAT AAAAAAGTTC GCTACTACAC AGTTAACAAT GATTCAATTA TGAGATTCAA 180
AAACGTGAAT ATTAGTTTCT TTAATACGAC ACACAGTATT CCTGATAGTT TAGGTGTCTG 240
TATTCACCCT TCATATGGTG CCATTGTGTA TACAGGTGAA TTTAAGTTTG ACCAAAGTTT 300
ACATGGACAT TATGCACCAG ATATTAAACG TATGGCAGAG ATTGGTGAAG AAGGCGTATT 360
TGTCTTAATC AGTGATTCTA CTGAGGCAGA GAAACCTGGA TATAATACTC CCGGAAAATG 420
TAATTGAACA TCATATGTAT GATGCCTTTG CCAAAGTGCG AGGTC 465

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

TTTAGAACAA TTAGATGCAC C 21

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

TCCGGGAGTA TTATATCCAG 20

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 527 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

GGCCCAAACC CATCCAAGTC CTTTTTAATT GACTTATTTA CATTATTTCT TTAATTTGGA 60
TTAACAAATT TTTTTCTATT TGANCCCTTT AATGTTNACT CCCCGTATCT AACAAGCAAG 120
TGATCATACT TCATTATTTT AGCAACTCCT TAATTTCCTC ATAAATGATG ATAAATATTT 180

CTTTAAACCT TGCTATATCT TCTTTAGTTG TAGTAGCCCC AAATGATAAT CTTATACTAC 240
CTTCAATAGA TTTGTCTGAT AATCCCATTG CAGCCAATAC TTCATTTAAT TTATTACGTT 300
TAGATGAACA AGCACTCGTC GTAGATATCA TAATGTCATA TTTTGAAAAA GCATTAACTA 360
ATACTTCACC TTTTACGCCA GGAAAACTAA GATTTAAAAC GAATGGTGAA CCTGAAGTTG 420
AAGAATTAAT ATAAACTCCA TGATATTTAT TTAAAAATTG ACGGACGTCA TTATTTAACT 480
CAGTAACAAA TGCATTCAAT GCTTCAAAGT TTTCATTAGC TCGTGCC 527

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

TTTTAGCAAC TCCTTAATTT CCTC 24

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

GCACGAGCTA ATGAAAACTT TG 22

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 811 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

GACAACTTGC TAAAGCACGT GATGAAAAAG TAAGTGAATA TGGAATTGAA CAAGCTGATG 60
GTACATTAAT TCAATATGAT AGTGAAGCCA AGATATATGA ACATTTTAAT GTGAATTTTA 120
TACCACCTGC TATGCGAGAA GATGGTAGCG AATTTGATAA AGATCTAAGT AATATCATTA 180
CATTAGATGA TATTAATGGT GATATTCATA TGCATACAAC GTATAGTGAT GGTGCGTTTT 240
CTATTCGAGA CATGGTAGAA GCAAATATCG CAAAAGGTTA TAAATTCATG GTAATTACTG 300
ATCATTCACA AAGTTTACGT GTTGCTAATG GCTTACAAGT GGAAAGACTT TTTANGACAA 360
AAACGAAGGA AATTAAGGCT TTAGATAAAG AATATAGTGA AATTGGATAT TTATTCAGGT 420
ACAAGAAATG GATATATTAA CCTGATGGCT CGCTGGATTA TGATGATGAA ATTTNAGCAC 480
AACTTGGATA TGTNATTGGA GCTATTCAAC AAAGCTTNAN CCAATCAGAA GAACAAATNA 540
TGGAACGGAT TAGCTAATGC ATGTCGCAAT CCATACGTGC GACATATAGC GCATCCAACA 600
GGGCGTATTA TAGGTAGAAG AGATGGTTAT AAACCGAATA TTGAACAATT AATGGCATTA 660
GCTGAAGAAA CGAATACAGT ATTAGAAATT AATGCCAATC CACATCGACT GGATCTTGAA 720
CGCTGAAATC GNTCGNNAAT ATCCAAATGT GAAATTAACT NTTAACACTG ATGGGCATCA 780
TNCAAATCAA TTNGATTTTN TGGAATTATG G 811

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

ACGTGATGAA AAAGTAAGTG

20

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

TCTTGTACCT GAATAAATAT CC

22

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 681 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

AGATCGTTCG CTAATTGACA ATTGATTAAA TCCCCTATTA CAAAATTGGA TATTACCTGT 60
TATATCTAAA AATCCACAAA TTGCTTTAGC AAGTGTTGAT NTGNCGGCAC CATTGTGACC 120
AACTATACTA AGCATTTCTC TTCTATAAAC ATTTAATTGA ACATTATTAA GTACACTATT 180
ACTATAGTCA CTATATTGAA CACATACCTC ATTTAATTCT AATAGCGGCN C. ATGTGTA 240
CTTATTATCA TTATGTGCAG ATGTNTCATC TATCCATTTN NNCACTTTAA NTTTAACATG 300
TTCACTCATA CAAACGACAC GTAANTTCGC TAAGTTATCA ATGGATTCGA CATCTACTTC 360
TGNATATTNA AGCGCTGNAC AGTATAATGG NACACGTATG CCTGCTTCTT TAAGCTTAGA 420
TGATTTTAGC AAATCACTAG GCGTTGTATT AGCGATGATT TTTCCATCTT TAAAAAGAAG 480
ANCTCTATCA AACGTATCAT CTAATGANTC TTCTAATCGA TGTTCGACAA TAATCATCGT 540

TGACTTTGTT TCTTCATGAA TATTGTNTAA CAATCTCAGC GTTTCATGTC CTGTCGCAGG 600
ATCTAAATTG GCCAGCGGCT CATCCAATAT TAAAATAGGC GTNCGATGGA TTAATATACC 660
ACCTAATGAA ACGCTCGTGC C 681

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

AATTGACAAT TGATTAAATC CCC 23

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

GCCAATTTAG ATCCTGCGAC 20

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 164 amino acids
(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Protien

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

Met Gly Met Val Ala Val Xaa Val Cys Thr Pro Pro Ile Gly Leu Gly
1 5 10 15
Leu Ala Thr Xaa Val Xaa Lys Tyr Lys Phe Asn His Ser Glu Arg Glu
20 25 30
Met Gly Lys Ala Xaa Phe Thr Met Gly Leu Phe Gly Ile Thr Glu Gly
35 40 45
Ala Ile Pro Phe Ala Ala Gln Asp Pro Leu Arg Ile Ile Pro Ala Asn
50 55 60
Ile Ile Gly Ala Met Ile Ala Ser Val Ile Ala Xaa Ile Gly Gly Val
65 70 75 80
Gly Asp Arg Val Ala His Gly Gly Pro Ile Val Ala Val Leu Gly Gly
85 90 95
Ile Asp His Val Leu Trp Phe Ile Phe Gly Xaa Ile Val Gly Ser Leu
100 105 110
Val Thr Met Pro Thr Val Leu Leu Xaa Arg Asn Thr Pro Val Ile
115 120 125
Ala Val Asp Ala Pro Ala Gln His Thr Gln Leu His Asp Thr Asp Ile
130 135 140
Thr Gln His Asp Thr Glu Val Asp Asn Val Asp Gly Thr Ser Glu Thr
145 150 155 160
Phe Thr Ser Gln

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 155 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

Met Asn Ile Glu Xaa Asp Ile Asn Gly Arg Pro Lys His Ile Tyr Ser
1 5 10 15
Ile Tyr Arg Xaa Met Met Lys Gln Lys Lys Gln Phe Asp Gln Ile Phe
20 25 30
Asp Leu Leu Ala Ile Arg Val Ile Val Asn Ser Ile Asn Asp Cys Tyr
35 40 45
Ala Ile Leu Gly Leu Val His Thr Leu Trp Lys Pro Met Pro Gly Arg
50 55 60
Phe Lys Asp Tyr Ile Ala Met Pro Lys Gln Asn Leu Tyr Gln Ser Leu
65 70 75 80
His Thr Thr Val Val Gly Pro Asn Gly Asp Pro Leu Glu Ile Gln Ile
85 90 95
Arg Thr Phe Asp Met His Glu Ile Ala Glu His Gly Val Ala Ala His
100 105 110
Trp Ala Tyr Lys Glu Gly Lys Lys Val Ser Glu Lys Asp Gln Thr Tyr
115 120 125
Gln Asn Lys Leu Asn Trp Leu Lys Glu Leu Ala Glu Ala Asp His Thr
130 135 140
Ser Ser Asp Ala Gln Glu Phe Met Glu Thr Leu
145 150 155

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 139 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

Asp Val Ala Lys Arg Leu Asn Ala Asn Ile Tyr Val Ser Gly Glu Gly
1 5 10 15
Glu Asp Ala Leu Gly Tyr Lys Asn Met Pro Ser Lys Thr Gln Phe Val
20 25 30
Lys His Gly Asp Ile Ile Gln Val Gly Asn Val Lys Leu Glu Val Leu
35 40 45
His Thr Pro Gly His Thr Pro Glu Ser Ile Ser Phe Leu Leu Thr Asp
50 55 60
Leu Gly Gly Gly Ser Xaa Val Pro Met Gly Leu Phe Ser Gly Asp Phe
65 70 75 80
Ile Xaa Xaa Gly Asp Ile Gly Arg Pro Asp Leu Leu Glu Lys Ser Cys
85 90 95
Ser Asn Lys Gly Phe Gly Thr Lys Leu Ala Arg Asn Lys Cys Met Ser
100 105 110
Pro Ile Lys Ile Leu Lys Ile Tyr Gln Thr Met Phe Lys Ser Gly Arg
115 120 125
Val Met Val Leu Glu Ala Leu Val Val Lys His
130 135

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 91 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

Met Tyr Gly Gly Val Thr Leu His Asp Asn Asn Arg Leu Thr Glu Glu
1 5 10 15
Lys Lys Val Pro Ile Asn Leu Trp Leu Asp Gly Lys Xaa Asn Thr Val
20 25 30
Pro Leu Glu Thr Val Lys Thr Asn Lys Lys Asn Val Thr Val Gln Glu

35 40 45
Leu Asp Leu Gln Ala Arg Arg Tyr Leu Gln Glu Lys Tyr Asn Leu Tyr
50 55 60
Asn Ser Asp Val Phe Asp Gly Lys Val Gln Arg Gly Leu Ile Val Phe
65 70 75 80
His Thr Ser Thr Glu Pro Ser Val Asn Tyr Asp
85 90

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 153 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

Met Leu Xaa Lys Met Leu Tyr Leu Leu Gln Ile His Gln Val Ile Pro
1 5 10 15
Ile Asn Ala Ile Ala Gln Ala Phe Asn Glu Lys Asp Gln Glu Arg Phe
20 25 30
Phe Gly Leu His Phe Phe Asn Pro Pro Arg Ile Met Xaa Leu Val Glu
35 40 45
Leu Ile Pro Thr Ser His Thr Lys Glu Ser Ile Ile Leu Asp Val Lys
50 55 60
Asn Phe Ala His Asn Val Leu Gly Lys Gly Val Ile Val Val Asn Asp
65 70 75 80
Val Pro Gly Phe Val Ala Asn Arg Val Gly Thr His Thr Met Asn Asp
85 90 95
Ile Leu Tyr Arg Ala Glu Gln His Lys Xaa Ser Xaa Val Asp Val Asp
100 105 110
Ala Leu Thr Gly Gln Ala Ile Gly Arg Pro Lys Thr Gly Thr Tyr Xaa
115 120 125
Leu Ser Asp Leu Val Gly Leu Xaa Ile Ala Xaa Ser Val Ile Lys Gly

130 135 140

Xaa Gln Xaa Val Pro Glu Glu Thr Pro

145 150

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 271 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

Met Lys His Leu Leu Gly Thr Lys Ser Gly Leu Leu Ala Thr Pro Asn

1 5 10 15

Glu Asp Glu Lys Pro Glu Glu Ile Thr Trp Arg Glu Glu Thr Thr Gly

20 25 30

Lys Leu Asp Leu Val Val Ser Leu Asp Phe Arg Met Thr Ala Thr Pro

35 40 45

Leu Tyr Ser Asp Ile Val Leu Pro Ala Ala Thr Trp Tyr Glu Lys His

50 55 60

Asp Leu Ser Ser Thr Asp Met His Pro Tyr Val His Pro Phe Asn Pro

65 70 75 80

Ala Ile Asp Pro Leu Trp Glu Ser Arg Ser Asp Trp Asp Ile Tyr Lys

85 90 95

Thr Leu Ala Lys Ala Phe Ser Glu Met Ala Lys Asp Tyr Leu Pro Gly

100 105 110

Thr Phe Lys Asp Val Val Thr Thr Pro Leu Ser His Asp Thr Lys Gln

115 120 125

Glu Ile Ser Thr Pro Tyr Gly Val Val Lys Asp Trp Ser Lys Gly Glu

130 135 140

Ile Glu Ala Val Pro Gly Arg Thr Met Pro Asn Phe Ala Ile Val Glu

145 150 155 160

Arg Asp Tyr Thr Lys Ile Tyr Asp Lys Tyr Val Thr Leu Gly Pro Val

165 170 175
Leu Glu Lys Gly Lys Val Gly Ala His Gly Val Ser Phe Gly Val Ser
180 185 190
Glu Gln Tyr Glu Glu Leu Lys Ser Met Leu Gly Thr Trp Ser Asp Thr
195 200 205
Asn Asp Asp Ser Val Arg Ala Asn Arg Pro Arg Ile Asp Thr Ala Arg
210 215 220
Asn Val Ala Asp Ala Ile Leu Ser Ile Ser Ser Ala Thr Asn Gly Lys
225 230 235 240
Leu Ser Gln Lys Ser Tyr Glu Asp Leu Glu Glu Gln Thr Gly Met Pro
245 250 255
Leu Lys Asp Ile Ser Ser Glu Arg Ala Ala Glu Lys Ile Arg Phe
260 265 270

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 143 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

Met Asp Ile Pro Asn Asn Leu Leu Ile Ala Gln Asn Val Pro Lys Glu
1 5 10 15
Ala Gln Met Ile Ala Glu Ala Gly Gln Ile Gly Ser Met Thr Val Ala
20 25 30
Thr Ser Met Ala Gly Arg Gly Thr Asp Ile Lys Leu Gly Glu Gly Val
35 40 45
Glu Ala Leu Ala Gly Leu Ala Val Ile Ile His Glu His Met Glu Asn
50 55 60
Ser Arg Val Asp Arg Gln Leu Arg Gly Arg Ser Gly Arg Gln Gly Asp
65 70 75 80
Pro Gly Ser Ser Cys Ile Tyr Ile Ser Leu Asp Asp Tyr Leu Xaa Lys

85 90 95
Arg Trp Ser Asp Ser Asn Leu Ala Glu Asn Asn Gln Leu Tyr Ser Xaa
100 105 110
Asp Ala Gln Arg Leu Ser Gln Ser Asn Leu Phe Asn Arg Lys Val Lys
115 120 125
Gln Ile Val Val Lys Ala Gln Arg Ile Ser Glu Arg Thr Arg Gly
130 135 140

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 221 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

Gly Glu Ser Ile Phe Val Gly Leu Ile Leu Gly Leu Gly Ile Gly Val
1 5 10 15
Leu Ala Gly Tyr Lys Pro Gly Asp Ile Ile Asn Leu Gly Met Ser Met
20 25 30
Ala Ala Val Met Val Leu Met Pro Arg Met Val Lys Ile Leu Met Glu
35 40 45
Gly Leu Met Pro Val Ser Glu Ser Ala Arg Thr Trp Leu Asn Lys Arg
50 55 60
Phe Gly Glu Arg Glu Ile Tyr Ile Gly Leu Asp Ala Ala Val Ala Leu
65 70 75 80
Gly His Pro Ala Val Ile Ser Thr Ala Leu Ile Leu Val Pro Ile Thr
85 90 95
Val Leu Leu Ala Val Ile Leu Pro Gly Asn Gln Val Leu Pro Phe Gly
100 105 110
Asp Leu Ala Thr Ile Pro Phe Val Val Ala Phe Ile Val Gly Ala Ala
115 120 125
Arg Gly Asn Ile Ile His Ser Val Ile Val Gly Thr Ile Met Ile Ala

130 135 140

Ile Ser Leu Tyr Ile Ala Thr Asp Val Ala Pro Ile Phe Thr Asp Met

145 150 155 160

Ala Lys Gly Thr Asn Val Gln Met Xaa Lys Gly Ser Ser Glu Xaa Ser

165 170 175

Ser Ile Asp Gln Gly Gly Asn Ile Xaa Asn Tyr Leu Ile Xaa Xaa Leu

180 185 190

Xaa Ser Leu Xaa Gln Xaa Lys Xaa Arg Xaa Val Cys Gly Gly Ser Phe

195 200 205

Ser Lys Asn Lys Arg Arg Thr Trp Gln Leu Arg Thr Ser

210 215 220

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 322 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

Met Tyr Gln Gln Ser Ser Phe Lys Glu Asn Leu Ile His Trp Phe Asp

1 5 10 15

Glu Asn Gln Arg Glu Met Pro Trp Arg Gln Thr Thr Asn Pro Tyr Tyr

20 25 30

Ile Trp Leu Ser Glu Val Met Leu Gln Gln Thr Gln Val Lys Thr Val

35 40 45

Ile Asp Tyr Tyr His Arg Phe Gly Xaa Arg Phe Pro Thr Val Glu Val

50 55 60

Leu Ser Gln Ala Ser Glu Asp Glu Val Leu Lys Tyr Trp Glu Gly Leu

65 70 75 80

Gly Tyr Tyr Ser Arg Ala Arg Asn Phe His Thr Ala Ile Lys Glu Val

85 90 95

Xaa Asp Lys Tyr Glu Gly Leu Val Pro Lys Asp Pro Asp Gln Phe Lys

```
                100                 105                 110
Ala Leu Lys Gly Val Gly Pro Tyr Thr Gln Ala Ala Val Met Ser Ile
        115                 120                 125
Ala Tyr Asn Val Pro Leu Ala Thr Val Asp Gly Asn Val Phe Arg Val
    130                 135                 140
Trp Ser Arg Leu Asn Asp Asp Tyr Arg Asp Ile Lys Leu Gln Ser Thr
145                 150                 155                 160
Arg Lys Ser Tyr Glu Gln Glu Leu Leu Pro Tyr Val Thr Thr Glu Ala
                165                 170                 175
Gly Thr Phe Asn Gln Ala Met Met Glu Leu Gly Ala Leu Ile Cys Xaa
            180                 185                 190
Pro Lys Asn Pro Leu Cys Leu Phe Xaa Pro Val Gln Glu Asn Cys Glu
        195                 200                 205
Ala Phe Asp Lys Gly Pro Phe Glu Lys Leu Pro Val Lys Ser Lys Asn
    210                 215                 220
Val Ser Lys Xaa Val Ile Glu Gln Ser Val Xaa Leu Ile Arg Asn Asn
225                 230                 235                 240
Gln Gly Gln Tyr Leu Leu Gln Lys Arg Arg Glu Xaa Leu Xaa Tyr Gly
                245                 250                 255
Met Trp Gln Xaa Pro Met Xaa Asp Ser Glu His Xaa Arg Arg Lys Met
            260                 265                 270
Xaa Glu Lys Ile Gly His Asp Ile Xaa Pro Xaa Glu Thr Pro Ile Xaa
        275                 280                 285
Glu Leu Thr His Gln Phe Thr His Leu Thr Trp Lys Ile Lys Val Tyr
    290                 295                 300
Ala Ala Ser Gly Ala Ile Asn Ile Xaa Thr Leu Pro Asp Asp Met Xaa
305                 310                 315                 320
Trp Val
```

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 151 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

Met Gly Ala Glu Asn Ile Ile Met Ala Phe Ile Leu Thr Arg Trp Ala
1 5 10 15
Trp Phe Cys Arg Val Ile Arg Thr Ser Val Met Gln Tyr Thr Ala Ser
20 25 30
Asp His Val Arg Phe Ala Lys Thr Ile Gly Met Asn Asp Met Lys Ile
35 40 45
Ile His Lys His Ile Met Pro Leu Thr Leu Ala Asp Ile Ala Ile Ile
50 55 60
Ser Ser Ser Ser Met Cys Ser Met Ile Leu Gln Ile Ser Gly Phe Ser
65 70 75 80
Phe Leu Gly Leu Gly Val Lys Ala Pro Thr Ala Glu Trp Gly Met Met
85 90 95
Leu Asn Glu Ala Arg Lys Val Met Phe Thr His Pro Glu Met Met Phe
100 105 110
Xaa Pro Gly Ile Ala Ile Gly Ile Ile Val Met Ala Phe Asn Phe Leu
115 120 125
Ser Asp Ala Leu Gln Asn Xaa Tyr Trp Ile Pro Arg Ile Ser Phe Leu
130 135 140
Lys Ile Asn Phe Arg Xaa Leu
145 150

(2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 221 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

Met Ile Phe Gly Lys Gly Thr Ala Lys Ala Thr Ser Tyr Gly Ala Gly
1 5 10 15
Ile Ile His Phe Leu Gly Gly Ile His Glu Ile Tyr Phe Pro Tyr Val
20 25 30
Leu Met Arg Pro Leu Leu Phe Ile Ala Val Ile Leu Gly Gly Met Thr
35 40 45
Gly Val Ala Thr Tyr Gln Ala Thr Gly Phe Gly Phe Lys Ser Pro Ala
50 55 60
Ser Pro Gly Ser Phe Ile Val Tyr Cys Leu Asn Ala Pro Arg Gly Glu
65 70 75 80
Phe Leu His Met Leu Leu Gly Val Phe Leu Ala Ala Leu Val Ser Phe
85 90 95
Val Val Ala Ala Leu Ile Met Lys Phe Thr Arg Glu Pro Lys Gln Asp
100 105 110
Leu Glu Ala Ala Thr Ala Gln Met Glu Asn Thr Lys Gly Lys Lys Ser
115 120 125
Ser Val Ala Ser Lys Leu Val Ser Ser Asp Lys Asn Val Asn Thr Glu
130 135 140
Glu Asn Ala Ser Gly Asn Val Ser Glu Thr Ser Ser Ser Asp Asp Asp
145 150 155 160
Pro Glu Ala Leu Leu Asp Asn Tyr Asn Thr Glu Asp Val Asp Ala His
165 170 175
Asn Tyr Asn Asn Ile Asn His Val Ile Phe Gly Cys Asp Ala Gly Met
180 185 190
Gly Ser Ser Ala Met Gly Ala Ser Met Leu Arg Asn Lys Phe Lys Lys
195 200 205
Ala Gly Ile Asn Asp Ile Thr Gly Tyr Lys Tyr Cys Asp
210 215 220

(2) INFORMATION FOR SEQ ID NO:90:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 227 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

Gly Thr Ser Val Ser Leu Gly Gly Ile Leu Ile His Arg Thr Pro Ile
1 5 10 15
Leu Ile Leu Asp Glu Pro Leu Ala Asn Leu Asp Pro Ala Thr Gly His
20 25 30
Glu Thr Leu Arg Leu Leu Xaa Asn Ile His Glu Glu Thr Lys Ser Thr
35 40 45
Met Ile Ile Val Glu His Arg Leu Glu Xaa Ser Leu Asp Asp Thr Phe
50 55 60
Asp Arg Xaa Leu Leu Phe Lys Asp Gly Lys Ile Ile Ala Asn Thr Thr
65 70 75 80
Pro Ser Asp Leu Leu Lys Ser Ser Lys Leu Lys Glu Ala Gly Ile Arg
85 90 95
Val Pro Leu Tyr Cys Xaa Ala Leu Xaa Tyr Xaa Glu Val Asp Val Glu
100 105 110
Ser Ile Asp Asn Leu Ala Xaa Leu Arg Val Val Cys Met Ser Glu His
115 120 125
Val Lys Xaa Lys Val Xaa Lys Trp Ile Asp Xaa Thr Ser Ala His Asn
130 135 140
Asp Asn Lys Tyr Thr Ser Xaa Pro Leu Leu Glu Leu Asn Glu Val Cys
145 150 155 160
Val Gln Tyr Ser Asp Tyr Ser Asn Ser Val Leu Asn Asn Val Gln Leu
165 170 175
Asn Val Tyr Arg Arg Glu Met Leu Ser Ile Val Gly His Asn Gly Ala
180 185 190
Xaa Xaa Ser Thr Leu Ala Lys Ala Ile Cys Gly Phe Leu Asp Ile Thr
195 200 205
Gly Asn Ile Gln Phe Cys Asn Arg Gly Phe Asn Gln Leu Ser Ile Ser
210 215 220
Glu Arg Ser
225

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

GCTCCTAAAA GGTTACTCCA CCGGC 25

What is claimed is:

1. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

(a) a polynucleotide having at least a 70% identity to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID Nos: 1,4,7,10,13,16,19,22,25 and 28;

(b) a polynucleotide which is complementary to the polynucleotide of (a); and

(c) a polynucleotide comprising at least 15 sequential bases of the polynucleotide of (a) or (b).

2. The polynucleotide of Claim 1 wherein the polynucleotide is DNA.

3. The polynucleotide of Claim 1 wherein the polynucleotide is RNA.

4. The polynucleotide of Claim 2 comprising the nucleotide sequence selected from the group consisting of SEQ ID Nos: 1,4,7,10,13,16,19,22,25 and 28.

5. An isolated polynucleotide comprising a member selected from the group consisting of:

(a) a polynucleotide having at least a 70% identity to a polynucleotide encoding the polypeptide expressed contained in NCIMB Deposit No. 40771 and selected from the group consisting of SEQ ID NOs: 1,4,7,10,13,16,19,22,25 and 28;

(b) a polynucleotide complementary to the polynucleotide of (a); and

(c) a polynucleotide comprising at least 15 bases of the polynucleotide of (a) or (b).

6. A vector comprising the DNA of Claim 2.

7. A host cell comprising the vector of Claim 6.

8. A process for producing a polypeptide comprising: expressing from the host cell of Claim 7 a polypeptide encoded by said DNA.

9. A process for producing a cell which expresses a polypeptide comprising transforming or transfecting the cell with the vector of Claim 6 such that the cell expresses the polypeptide encoded by the cDNA contained in the vector.

10. A process for producing a polypeptide of the invention or fragment comprising culturing a host of claim 7 under conditions sufficient for the production of said polypeptide or fragment.

11. A polypeptide comprising an amino acid sequence selected from the group consisting essentially of: 79,80,81,82,83,84,85,86,87 and 88.

12. An antibody against the polypeptide of claim 11.

13. An antagonist which inhibits the activity of the polypeptide of claim 11.

14. A method for the treatment of an individual having need of a polypeptide of the invention comprising: administering to the individual a therapeutically effective amount of the polypeptide of claim 11.

15. The method of Claim 14 wherein said therapeutically effective amount of the polypeptide is administered by providing to the individual DNA encoding said polypeptide and expressing said polypeptide *in vivo*.

16. A method for the treatment of an individual having need to inhibit a polypeptide of the invention comprising: administering to the individual a therapeutically effective amount of the antagonist of Claim 13.

17. A process for diagnosing a disease related to expression of the polypeptide of claim 11 comprising:

determining a nucleic acid sequence encoding said polypeptide.

18. A diagnostic process comprising: analyzing for the presence of the polypeptide of claim 11 in a sample derived from a host.

19. A method for identifying compounds which bind to and inhibit an activity of the polypeptide of claim 11 comprising:

contacting a cell expressing on the surface thereof a binding for the polypeptide, said binding being associated with a second component capable of providing a detectable signal in response to the binding of a compound to said binding, with a compound to be screened under conditions to permit binding to the binding; and

determining whether the compound binds to and activates or inhibits the binding by detecting the presence or absence of a signal generated from the interaction of the compound with the binding.

20. A method for inducing an immunological response in a mammal which comprises inoculating the mammal with a polypeptide of the invention, or a fragment or variant thereof, adequate to produce antibody to protect said animal from disease.

21. A method of inducing immunological response in a mammal which comprises, through gene therapy, delivering gene encoding a fragment of a polypeptide of the invention or a variant thereof, for expressing such polypeptide, or a fragment or a variant thereof *in vivo* in order to induce an immunological response to produce antibody to protect said animal from disease.

22. An immunological composition comprising a DNA which codes for and expresses a polynucleotide of the invention or protein coded therefrom which, when introduced into a mammal, induces an immunological response in the mammal to a given such polynucleotide or protein coded therefrom.

23. A polynucleotide consisting essentially of a DNA sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence of the invention under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence or a fragment thereof; and isolating said DNA sequence.